Investigating different rational design approaches to increase brightness in red fluorescent proteins

Sandrine Legault

Thesis submitted in partial fulfillment of the requirements for

the degree of Master of Science in Chemistry

University of Ottawa

Ottawa, Ontario, Canada

September 2021

Abstract

Red fluorescent proteins (RFPs) are used extensively in biological research because their longer emission wavelengths are less phototoxic and allow deeper imaging of animal tissue. However, far-red RFPs generally display low brightness, emphasizing the need to develop brighter variants. Here, we investigate three approaches to rigidify the RFP chromophore to increase the quantum yield, and thereby brightness. We first used computational protein design on a maturation-efficient mRojo-VHSV variant previously engineered in our lab to introduce a *Superdecker* motif, a parallel pi-stack comprising aromatic residue side chains and the phenolate moiety of the chromophore, which we hypothesized would enhance chromophore packing and reduce non-radiative decay. The best mutants identified showed up to 1.7-fold higher quantum yield at pH 9, relative to their parent protein. We next postulated that brightness could be further increased by rigidifying the chromophore via branched aliphatic residues. Computational protein design was performed on a dim mCherry variant, mRojoA, followed by directed evolution on the brightest mutant. The combination of these methodologies yielded mSandy2, the brightest *Discosoma*-derived monomeric RFP with an emission maximum above 600 nm. Finally, we aimed to increase brightness by focusing on positions where residue rigidity correlated to quantum yield in mCherry-related RFPs according to NMR data that had been previously acquired in our lab. Combinatorial site-saturation mutagenesis was performed on two different surface patches of mCherry at positions 144/145/198 and 194/196/220. Our results demonstrated that surface residues may not be adequate targets for this approach. Altogether, the work herein presents unique rational design methodologies that can be used to increase brightness in RFPs.

Acknowledgements

Before we dive into the scientific outlook of my graduate studies, I would like to take the time to thank several people that have made this journey truly memorable.

First and foremost, I would like to thank my supervisor, Dr. Roberto Chica. I could not have asked for a better mentor. Because of him, I had the opportunity to get out of my comfort zone, to attend many conferences and to build a solid network. Most importantly, he encouraged me to reach for my goals, no matter how far they seemed at the time. Without him, I would not be the scientist I am today and for that I am forever grateful.

Secondly, I would like to express my gratitude towards one of the most respectable researchers I have had the chance to meet through conferences: Dre Joëlle Pelletier. To me, she is an amazing role model for all young women in chemistry. It is always a pleasure to talk about research updates with her. Moreover, she has always kindly took the time to help me in my academic journey and I will never take this for granted.

Thirdly, I probably would not have even considered laboratory research if not for Dr. Matthew Eason and Dr. Adam Damry who have been guiding me since my undergraduate years. It has been a privilege to have worked alongside such remarkable scientists. Thank you for sharing with me your knowledge and your precious advice.

I would also like to thank my colleagues – Chica Lab members and others – who have supported me and have made my graduate studies such a great experience. Thank you for the heart-felt conversations and the amazing friendships.

Last but certainly not least, my most sincere thanks go to Lynne Vinet and Yves Legault. I find myself truly blessed to have had the best parents and role models. They have been there for me every step of the way, through the highest of highs and the lowest of lows. Words cannot express my gratitude. If I am the compassionate and capable person I am today, it is because of you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

FP – Fluorescent Protein

GFP – Green Fluorescent Protein

FRET – Förster Resonance Energy Transfer

RFP – Red Fluorescent Protein

PDB – Protein Data Bank

EC – Extinction Coefficient

QY – Quantum Yield

TICT – Twisted Intramolecular Charge Transfer

PCR – Polymerase Chain Reaction

FACS – Fluorescence-Activated Cell Sorting

CPD – Computational Protein Design

MSD – Multi State Design

FPLC – Fast Protein Liquid Chromatography

LB – Luria-Bertani or Lysogeny Broth

IPTG – Isopropyl β-D-1-thiogalactopyranoside

PBS – Phosphate Buffered Saline

Chapter 1: Introduction

1.1. Origin and Applications of Red Fluorescent Proteins

The human body is composed of more than ten trillion cells[1] that work in harmony to ensure proper bodily function and survival. However, in some cases, abnormalities may arise from mutations in the genetic code occasionally leading to cellular malfunction or dysregulation which give rise to disease. In order to come up with cures or therapies, scientists must identify the misregulated elements responsible for the disease phenotype. One of many important research tools for this purpose is fluorescence-based cellular imaging. In particular, fluorescence microscopy techniques that rely on fluorescent proteins (FPs) allow us to detect and analyze cells, organelles or molecules with different temporal and spatial resolutions and play an important role in drug discovery, for example to screen for cancer therapeutics[2, 3].

The first isolated FP, the green fluorescent protein (GFP), was discovered in 1962 from the *Aequorea victoria* jellyfish[4], was cloned in 1994 in bacterial and prokaryotic cells from isolated complementary DNA, and was highlighted as a valuable cofactor-independent, genetically-encoded imaging tool to monitor gene expression and track proteins in living organisms[5]. Soon after, Tsien *et al.* (1996) provided insight on the chromophore fluorescence, reported the first engineering of FPs to increase brightness, expanded the color palette by generating the blue fluorescent protein (BFP), and highlighted the benefit of FP engineering for multiplex imaging as well as Förster resonance energy transfer (FRET) between FPs that possess overlap between emission excitation spectra[6]. The great impact of these initial findings has been widely recognized, and the Nobel Prize in Chemistry 2008 was awarded to Osamu Shimomura, Martin Chalfie

and Roger Y. Tsien for their respective contributions described above[7, 8]. Since then, scientists have identified and engineered a variety of FPs with emission wavelengths that span the entire visible spectrum[9]. One example of these was the engineering of *Aequorea*-based red fluorescent proteins (RFPs) using combinatorial site-saturation mutagenesis followed by several rounds of random mutagenesis[10]. However, these efforts have been unsuccessful as only a small portion (1-3%) of the proteins matured to the red chromophore.

One important finding following the isolation of the *Aequorea Victoria* GFP was the discovery of the first isolated RFP DsRed in 1999, a tetrameric protein from the red mushroom coral *Discosoma sp.* shown to be useful for *in vivo* labeling of *Xenopus* embryos[11]. Other RFPs were isolated from the bubble-tip anemone *Entacmaea quadricolor*: the tetrameric eqFP611[12] and the dimeric eqFP578[13]. However, the oligomeric state of these RFPs are undesirable for multiple applications as their tendency to form oligomers in solution may favor aggregation and negatively impact imaging results. Moreover, monomeric RFPs and thereby smaller fusion constructs better facilitate experimental manipulation. Engineering of DsRed, eqFP578 and eqFP611 gave rise to various useful monomeric RFPs (**Table 1.1**). For example, creation of a monomeric variant of DsRed yielded mRFP1[14] which provided a template for the engineering of a series of yellow to far-red FPs known as the family of mFruits[15, 16], from which emanated the most widely used RFP mCherry[15]. In turn, engineering of eqFP578 and eqFP611 respectively yielded the bright RFPs mKate[17] and mRuby[18], which were subsequently further developed into brighter versions (mKate2[19], mRuby2[20], and mRuby3[21]) and into the most red-shifted but dim RFPs mNeptune684[22] and mGarnet2[23], respectively. Finally, the brightest RFP to

date, mScarlet, was evolved using a synthetic gene template containing previously identified beneficial mutations from various engineered RFPs including those mentioned above[24].

Table 1.1. Spectroscopic characteristics of various representative red fluorescent proteins. This table compares the spectroscopic characteristics of bright RFPs originating from the three different wild-type parent proteins available DsRed, eqFP578 and eqFP611, as well as the brightest RFP mScarlet engineered from a synthetic gene template.

Protein	λ_{ex} (nm)	λ_{em} (nm)	Φ_F	ε ($mM^{-1}cm^{-1}$)	Brightness ($mM^{-1}cm^{-1}$)	Ref.
Origin: Discosoma sp.						
DsRed	558	583	0.68	72.5	49.30	11
mRFP1	584	607	0.25	50	12.50	14
mCherry	587	610	0.22	72	15.84	15
Origin: Entacmaea quadricolor						
eqFP578	552	578	0.54	102	55.08	13
mKate	588	635	0.33	45	14.85	17
mKate2	588	633	0.40	62.5	25.00	19
eqFP611	559	611	0.45	78	35.10	12
mRuby	558	605	0.35	112	39.20	18
Origin: Synthetic gene template						
mScarlet	569	594	0.70	100	70.00	24

λ_{ex} and λ_{em} correspond to the excitation and emission maxima.
The brightness is the product of the quantum yield (Φ_F) and the extinction coefficient (ε).
Data taken from the Fluorescence Protein Database, fpbase.org.

Engineered RFPs can be used as acceptors in FRET-based cellular probes[25], as fusion constructs to track proteins within cells[24, 26], and as genetically-encoded biosensors for detection of signal transduction[27] or ion sensing[28, 29]. Furthermore, RFPs possess advantages relative to other FPs and therefore have been of particular interest in the field. Primarily, their longer excitation wavelengths ($\geq$ 558 nm[11]) allow for reduced cellular phototoxicity due to their low-energy fluorescence, as described by the inverse correlation between energy and wavelength in the following equation:

$$E = hc/\lambda \qquad \text{eq. [1]}$$

where E is the energy, h is Planck's constant, c is the speed of light, and λ is the wavelength. In addition, their red-shifted excitation wavelengths allows for deeper tissue penetration[30] and lower background auto-fluorescence, which result from reduced tissue light scattering and tissue absorption[31]. Because of this, RFPs are not only useful for less phototoxic *in cellulo* imaging purposes but are also advantageous for non-invasive *in vivo* imaging of deep tissue in whole-animal research models[17, 32, 33]. For all applications, bright and red-shifted RFPs are hence desirable. However, as evidenced in **Figure 1.1.**, red-shifted monomeric RFPs generally display low brightness relative to other widely used green and yellow monomeric FPs, as well as often lower brightness than their natural parent proteins (as shown in **Table 1.1**). Moreover, as the wavelength is increased, RFP brightness tends to decrease substantially. This observation can be explained by the "energy gap law" discussed in section 1.3 below, and emphasizes the need to develop brighter variants. The work herein tackles this problem by investigating different rational design approaches to increase RFP brightness.

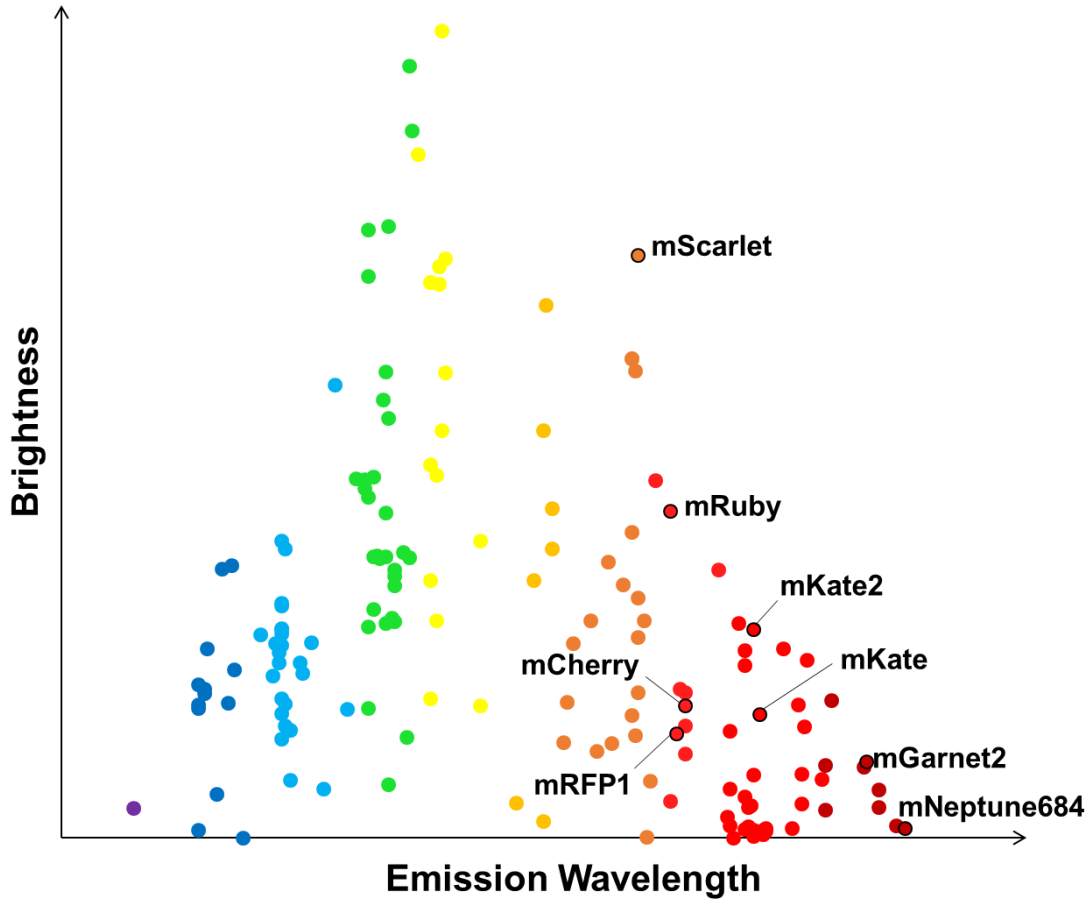


Figure 1.1. Red-shifted fluorescent proteins generally display low brightness relative to other fluorescent proteins. Overall, RFPs display lower brightness relative to other green and yellow FPs. Furthermore, RFP brightness tends to be inversely correlated with emission wavelength. Each dot represents a monomeric fluorescent protein color-coded with respect to its emission wavelength. (Data used to create this figure was taken from the Fluorescent Protein Database, fpbase.org.)

1.2. Structural Overview of RFPs

Despite originating from different organisms, the RFPs discussed in this work are members of the GFP-like superfamily of FPs who all share a similar fold. They are composed of a β-barrel consisting of 11 β-strands surrounding an α-helix which contains the chromophore (4-(p-hydroxybenzylidene)-5-imidazolinone) responsible for fluorescence (**Figure 1.2**). In RFPs, the chromophore is autogenically formed by the

cyclization of the Xaa66-Tyr67-Gly68 tripeptide in the protein core (where Xaa is any amino acid at position 66, and the numbering is based on the sequence of mCherry), catalyzed via Glu215 and Arg96 catalytic residues[34, 35]. The main distinction of the RFP chromophore is the additional oxidation step in the chromophore maturation pathway which leads to the formation of an *N*-acylimine group on the peptide backbone, responsible for the extended conjugation and the red-shifted fluorescence (**Figure 1.2 b**).

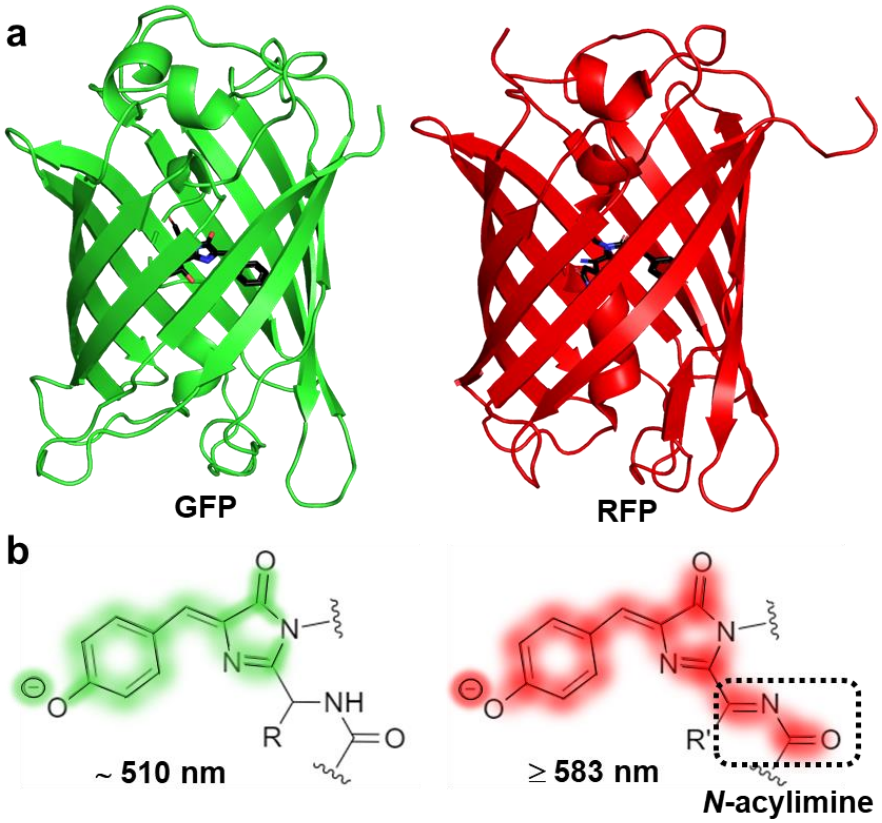


Figure 1.2. Structural characteristics of FP monomers. (a) Structure overview of the similar folds of GFPs (GFP, PDB ID: 1EMA[36]) and RFPs (mCherry, PDB ID: 2H5Q[37]), shown in green and red cartoons respectively. On each structure, the buried chromophore is shown in black sticks. (b) Comparison of the green and red chromophores. The conjugated systems responsible for the emission wavelengths (λem = 510 nm; λem ≥ 583 nm) are highlighted in green and red, respectively. The extended conjugated system in the RFP chromophore is caused by the presence of the *N*-acylimine group, circled with a dotted line.

As described by Strack *et al.* (2010), the formation of "the red chromophore [...] occurs by a branched pathway"[34] (**Figure 1.3**). Similarly to the GFP chromophore maturation[38], initial steps involve the backbone cyclization of the tripeptide via nucleophilic attack of the nitrogen amide of Gly68 on the carbonyl of the Xaa66, forming a heterocycle which then gets oxidized. These steps are facilitated by the catalytic residue Arg96 side chain whose guanidinium group forms a hydrogen bond with the oxygen of the Tyr67 carbonyl thereby lowering the pKa value of the Gly68 nitrogen[39] and by the catalytic residue Glu215 which plays the role of a general base facilitating deprotonation of the Gly67 amide nitrogen[38]. The resulting hydroxylated cyclic imine intermediate can either irreversibly undergo dehydration and reversible deprotonation to form the green chromophore species which absorb around 390 nm and 510 nm respectively or, can reversibly lose a hydroxyl moiety, constituting the "branch point" in the pathway[35] (**Figure 1.3, step 5**). This branching point represents a kinetic competition between formation of the green species and formation of a blue intermediate towards the red chromophore. At this point, if O_2 access to the chromophore is not restricted and red chromophore maturation is efficient, the reaction is shifted in the direction of the loss of a hydroxyl moiety and the irreversible oxidation step forming the *N*-acylimine on the peptide backbone and the blue species. This intermediate is further hydroxylated and dehydrated, and a final deprotonation yields the fluorescent red chromophore which absorbs around 590 nm in the case of most DsRed-derived RFPs known as the mFruit family, such as mCherry.

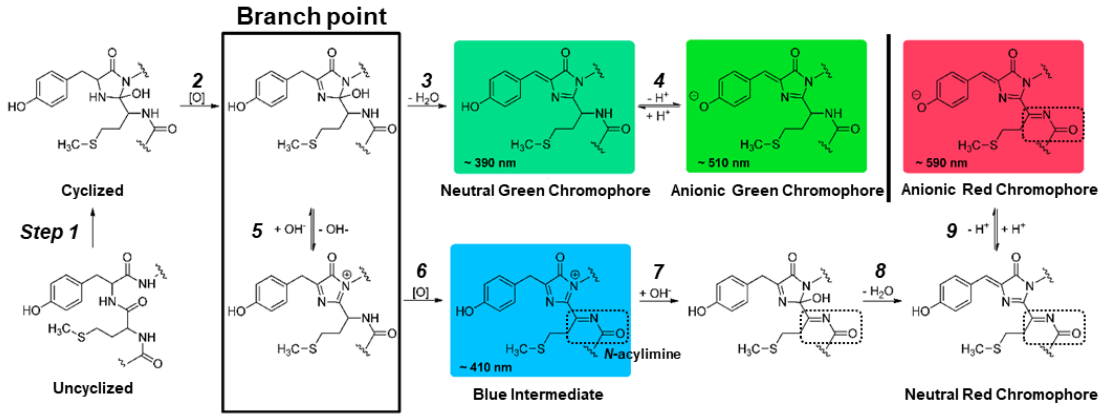


Figure 1.3. Branched chromophore maturation pathway in RFPs. Example of the mCherry chromophore formed by the Met66-Tyr67-Gly68 tripeptide. Chromophore maturation follows these reaction steps: cyclization (*step 1*) of the tripeptide, oxidation (*2*), dehydration (*3*) forming the neutral green chromophore, and reversible deprotonation (*4*) forming the anionic green chromophore. After the first oxidation step (*2*), efficient red chromophore maturation leads to the loss of a hydroxyl group (branch point, *5*) on the path to a second irreversible oxidation step (*6*) to form the *N*-acylimine group (circled with a dotted line) and the blue intermediate which then undergoes hydroxylation (*7*) and dehydration (*8*) to form the neutral red chromophore, further deprotonated to form the fluorescent anionic red chromophore. (Figure adapted from Moore *et al*. (2012)[35].)

There can be competing requirements between the chromophore maturation and other RFP properties, complicating their engineering. For example, O_2 access is critical for chromophore maturation as it is required for two steps in the formation pathway. However, too much O_2 access has also been shown to cause faster photobleaching[40]. In addition, engineering of a tighter chromophore environment has often been exploited to enhance RFP brightness, as will be discussed below. On the other hand, a too compact core filled with rigid bulky residues may result in maturation deficiency[41] due to restriction of the motions required by the peptide backbone to undergo the initial cyclization step in the chromophore maturation pathway. For these reasons, improving one property can often impact others and needs to be taken into account when engineering RFPs depending on their desired end application. However, in the case of this thesis, we will interrogate

different methodologies to exclusively improve RFP brightness and will therefore mainly focus on our understanding of how fluorescence can be enhanced.

1.3. Fluorescence and Brightness in RFPs

The RFP chromophore is responsible for fluorescence and its properties are directly linked to brightness. The brightness is the product of the extinction coefficient (ε, $M^{-1}cm^{1}$) which is the ability of the chromophore to absorb light and the quantum yield of fluorescence (Φ_F) which is its ability to convert the light absorbed into fluorescence emission, as depicted by the equations below:

$$\text{Brightness} = \varepsilon \cdot \Phi_F \qquad \text{eq. [2]}$$

$$\text{Beer-Lambert's Law: } \varepsilon = \frac{A}{C \cdot l} \qquad \text{eq. [3]}$$

where *A* is the absorbance of the chromophore at a particular wavelength, *C* is the molar concentration of chromophore in solution (in M), and *l* is the length of the path traveled by light in solution (in cm).

$$\Phi_F = \frac{\text{number of photons emitted}}{\text{number of photons absorbed}} \qquad \text{eq. [4]}$$

where the quantum yield value thereby ranges between 0 and 1.

The quantum yield can also be calculated using the equation below which specifies that it is the proportion of fluorescence relative to the sum of all other processes undergone by the excited molecule as shown in **Figure 1.4**:

$$\Phi_F = \frac{k_F}{k_F + \Sigma\, k_{non-radiative}} = \frac{k_F}{k_F + k_{IC} + k_{EC} + k_{ISC} + k_{Di} + k_{PDi}} \qquad \text{eq. [5]}$$

where k_F is the rate constant of fluorescence, k_{IC} is the rate constant of internal conversion (from a higher electronic state to a high vibrational level of a lower electronic state without photodecomposition), k_{EC} of external conversion (energy transfer with the solvent) , k_{ISC} of intersystem crossing (crossover between electronic states with different multiplicity), k_{Di} of dissociation (bond rupture resulting from excitation to a high enough vibrational level), and k_{PDi} of pre-dissociation (internal conversion that occurs prior to bond rupture)[42-44]. In eq. [5], k_F represents the radiative pathway of interest and all other rate constants correspond to non-radiative pathways.

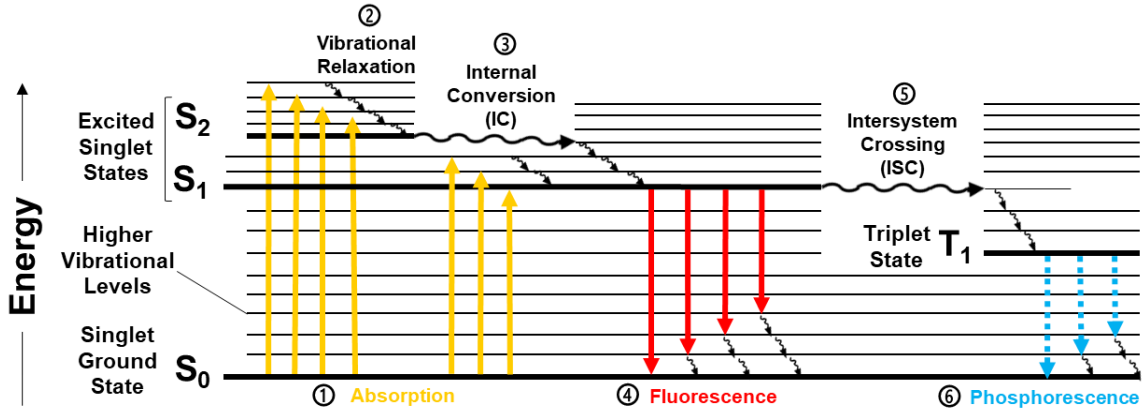


Figure 1.4. Jablonski diagram illustrating the different processes undertaken by an excited molecule. Irradiation of the molecule can excite an electron from the ground state to an excited singlet state (**1**). When it is excited to higher vibrational levels of a higher order excited singlet state (S_{2+n}), the molecule undergoes vibrational relaxation to the lowest vibrational level of that electronic state (**2**). It can then undergo internal conversion (IC) to reach the first excited singlet state (**3**). From S_1, the molecule can either release the absorbed energy via fluorescence (**4**), or internal conversion, to the ground state S_0. From S_1, it can also undergo intersystem crossing (ISC) to the triplet state T_1 (**5**) from which, phosphorescence can be emitted (**6**). Note: ISC (µs timescale) is generally slow relative to fluorescence lifetime (ns timescale) and is not usually observed in RFPs[45]. Color-labeled processes are radiative means, whereas black-labeled processes represent non-radiative decay.

In the case of the RFP chromophore, the extinction coefficient (also termed molar absorptivity) translates to the ability of the red chromophore to absorb light at its absorption maximum (around 580–590 nm for most RFPs). Since this value depends on the concentration of red chromophore in a RFP population, it is partially influenced by the efficiency of red chromophore maturation described earlier. Most RFPs display extinction coefficient values ranging from ~ 30 mM^{-1} cm^{-1} to ~ 100 mM^{-1} cm^{-1} which is corresponds to what is observed in other FPs[46-48]. In addition, it is not well-known how structural changes to the protein structure can beneficially impact the extinction coefficient which is why most efforts to increase it have been made using random mutagenesis[24] or site-saturation mutagenesis of residues surrounding the chromophore[49].

On the other hand, the quantum yield, which is the ability of the chromophore to convert the light absorbed into fluorescence, tends to be lower in monomeric RFPs with an emission maximum above 600 nm (0.02 to 0.47, mRojoA[50] and cgfmKate2[51]) compared to that of monomeric yellow or green FPs (0.12 to more than 0.80, mHoneydew[15] and mNeonGreen[50]). As shown by eq. [5], the quantum yield can be improved by disfavouring non-radiative decay pathways. There are two major pathways by which the red chromophore can undergo non-radiative decay and limit fluorescence emission: the energy gap law and twisted intramolecular charge transfer[52] (TICT). The energy gap law is true for most red-shifted molecules and, in this case, reflects the facilitated non-radiative relaxation pathway that the red chromophore can undergo due to overlap between the ground and excited states. In brief, the RFP chromophore's longer excitation and emission wavelengths result from a lower energy difference (as described earlier in this chapter by the eq. [1]) between the ground and excited electronic states, $S_0 \rightarrow S_1$. Because of these

"small energy gaps", there can be overlap between the lowest vibrational level of S_1 and the higher vibrational levels of S_0. This leads to faster internal conversion and vibrational relaxation to the ground state S_0[53, 54]. This explains why the quantum yield tends to decrease with the longer wavelengths. On the other hand, TICT is not dependent of the wavelength but is rather influenced by conformational flexibility of the chromophore known as "hula-twisting" of its phenolate moiety, as described by Baffour-Awuah and Zimmer (2004) in the case of the GFP chromophore[55]. As illustrated in **Figure 1.5**, this motion corresponds to the rotation of the phenol ring, or torsion of the dihedral angle γ defined by atoms CA-CB-CG-CD, and to the torsion of the dihedral angle β defined by N1-CA-CB-CG due to isomerization[56]. According to Drobizhev *et al.* (2021), twisting of the RFP chromophore is predominantly due to torsions of the dihedral γ and leads to an intersection of the excited and ground states' potential energy surfaces and charge transfer across the chromophore, thereby leading to rapid non-radiative loss of energy[52].

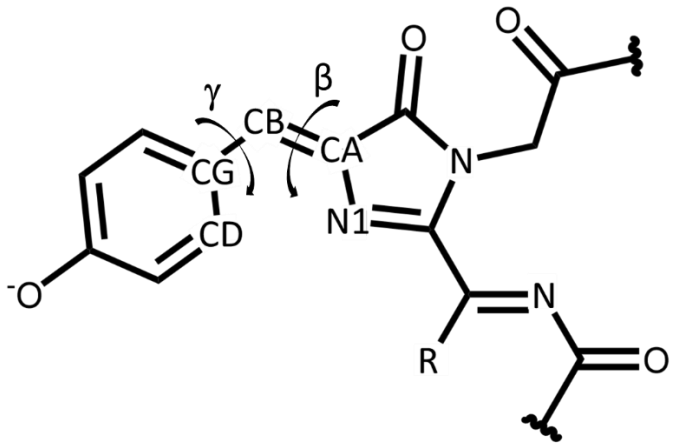


Figure 1.5. Hula-twist motion in the RFP chromophore. The twisted state of the RFP chromophore responsible for non-radiative decay is caused by torsions of the dihedral angles we name β and γ. Dihedral β is defined by the atoms N1-CA-CB-CG, and dihedral γ is defined by CA-CB-CG-CD, resulting in a dim non-planar chromophore.

Altogether, the more flexible the RFP chromophore, the more its energy will be released via non-radiative TICT resulting in low quantum yield. In contrast, the more rigid the chromophore, the less likely it will undergo non-radiative decay and so the more its energy will be released via fluorescence, resulting in higher quantum yield (**Figure 1.6)**. Because reduced conformational dynamics of the chromophore phenolate moiety can be engineered via enhanced packing, this determinant of the quantum yield is an attractive target for engineering RFP brightness and has been previously exploited in the field, as will be discussed in the section below.

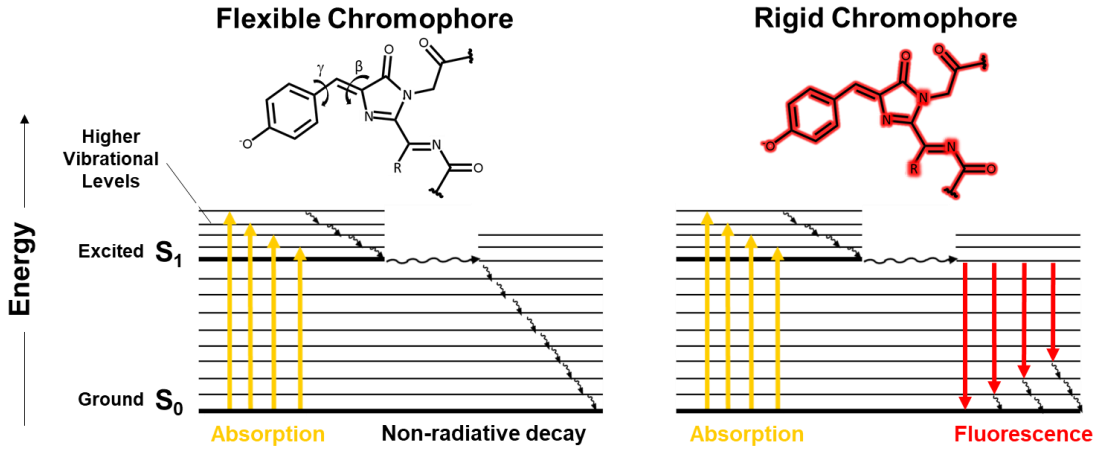


Figure 1.6. Major relaxation pathways undertaken by flexible or rigid chromophores. In the case of a flexible chromophore (specifically, torsions of dihedral angles β and γ), the energy absorbed after irradiation to the excited electronic state (yellow arrows) will mainly be released via non-radiative decay represented by vibrational relaxation between vibrational levels (small black arrows). Whereas, a rigid chromophore will most likely release the energy via radiative means, in this case fluorescence (red arrows).

1.4. Previously Reported Engineering of Brightness in RFPs

The problem of low RFP brightness has been addressed in the past and various successful case studies of bright RFP engineering have been previously reported, using two main approaches: directed evolution and rational design.

Directed evolution

Directed evolution has recently been the talk of the field because of its great impact on the manufacturing of biofuels and pharmaceuticals in industry, leading to the award of the Nobel Prize in Chemistry 2018 to Frances Arnold, George Smith and Sir Gregory Winter for its use in evolving enzymes, peptides and antibodies[57-59]. This methodology relies on random mutagenesis of the gene of interest to yield large libraries of mutant genes (often > 10^3 sequences) and is made to mimic natural selection. More specifically, it mainly depends on error-prone polymerase-chain reaction (PCR) by a low fidelity polymerase, albeit some also use DNA repair[16], followed by screening for enhanced properties of interest which should be amenable to high-throughput screening methods. For example, variants with increased quantum yields can be identified by screening for enhanced brightness using fluorescence-activated cell sorting (FACS) since quantum yield and brightness are correlated. Many iterative rounds of random mutagenesis and screening are usually required to increase the amount of beneficial mutations and obtain significantly improved variants. Directed evolution is an attractive approach because it has the advantage of finding unpredictable beneficial mutations, some of which are distal from the active site of the protein or in our case outside of the first shell of the RFP chromophore,

without *a priori* knowledge of the protein structure and has hence been widely used to engineer bright RFPs with various spectroscopic characteristics.

For example, most if not all RFPs from the mFruits superfamily have been evolved using this approach. Shaner *et al.* (2004) subjected mRFP1.5, an evolved variant of mRFP1, to multiple rounds of directed evolution to produce two brighter RFPs: mStrawberry and mCherry[15]. These variants are 2- and 1.3-fold brighter than mRFP1 respectively and mStrawberry displays an absolute 0.04 increase in quantum yield. Another case study that had a great impact in the field is the engineering of mScarlet, the brightest monomeric RFP with a high quantum yield of 0.70. mScarlet was identified after several rounds of random mutagenesis and screening for enhanced brightness and efficient maturation starting from a synthetic template which was designed based on multiple beneficial mutations that had been previously identified in different RFPs to be responsible for efficient red chromophore maturation, monomerization and high quantum yields[24]. Although the benefit of directed evolution in identifying unpredictable beneficial mutations has been particularly useful in these aforementioned case studies, it remains expensive in terms of resources, time and effort since large amounts of mutants ranging from thousands to millions need to be tested.

Rational design

In contrast, rational design is an alternate approach that requires experimental testing of fewer mutants ($\leq 10^2$) as the library is focused using design hypotheses that base themselves on knowledge of the protein structure. For example, mRuby3 was identified using combinatorial mutagenesis at specific positions on the protein surface of mRuby2 to improve protein folding and positions near the chromophore that the authors hypothesized

would enhance "packing of residues near the chromophore [to] reduce excited-state vibrations that could lead to non-radiative decay"[21], similarly to what we discussed earlier. Using this method, they were able to increase the brightness of mRuby2 by more than 1.3-fold and the quantum yield by an absolute value of 0.07. Another interesting rational design strategy was performed by Pandelieva *et al.* (2016) and will serve as the basis of our hypotheses in the next two chapters of this thesis. In this study, the authors rigidified the chromophore by sandwiching the chromophore phenolate moiety between two aromatic residues and introducing additional beneficial first shell mutations, which increased the quantum yield and the brightness of mRojoA, a dim red-shifted mCherry variant[50], by 2-fold[60]. Rational design is especially useful when crystal structures of the FPs of interest are available since the introduction of beneficial mutations require knowledge of the system. However, this can also serve as a limitation to this approach since not all RFPs are available in the Protein Data Bank (PDB)[61]. Moreover, using this type of rational design can encounter epistasis issues where beneficial mutations on one protein backbone can be detrimental on another.

Computational protein design

Another rational design methodology is computational protein design (CPD). CPD enables the design of amino-acid sequences that can stabilize a target protein scaffold and create a novel function. It can be used to predict whether different combinations of residues would result in stable protein folds, thereby yielding stable sequences to test experimentally. CPD relies on optimization algorithms that can efficiently search through large protein sequence spaces for combinations of amino-acid residues that result in stable mutants. The method requires a backbone template (structure obtained by x-ray

crystallography or protein NMR) on which residues of interest are designed at specific positions by side-chain placement. Each of these residues is represented by a collection of discrete side-chain rotamers, named rotamer library, which correspond to low energy conformations of amino-acid side chains observed in known protein structures. This step allows the sampling of different side chain conformations during the simulation to find the combinations that result in the lowest energy on the chosen backbone. The term "energy" mentioned here is the calculated output of an energy function that contains non-bonded interaction (for example electrostatic and Van der Waals interactions) and solvation energy terms used to compute a potential energy value. This potential energy value reflects the predicted stability for different sequences on the backbone template, where the best ranked sequences (predicted to be most stable) are those with the lowest predicted potential energy. In contrast with the rational design approaches discussed in the previous paragraph, CPD has the advantage of predicting combinations of mutations that should be non-detrimental to protein stability and is therefore a useful "pre-screening" technique for eliminating mutant proteins that would not be stable.

CPD was performed by Chica *et al.* (2010) on the mCherry scaffold on which thirteen residues with side chains pointing towards the chromophore were designed in order to identify combinations of mutations that together could interact with the chromophore to red-shift its emission wavelength without hindering protein stability[50]. Using this approach, the authors identified mRojoA, a variant displaying a 22-nm bathochromic shift relative to mCherry. CPD simulations have also been used to create monomeric RFPs as shown by Wannier *et al.* (2015, 2018)[62, 63]. The authors performed CPD on a tetrameric RFP on which they targeted surface residues at positions found on the protein's oligomeric interfaces.

Evaluation of the 96 designed sequences with the lowest computed energy yielded 93 variants with the desired monomeric structure[62]. These past results highlight the utility of CPD as a powerful *in silico* pre-screening strategy. Although this methodology has been useful for engineering red-shifted or monomeric RFPs, it has not, to the best of our knowledge, been used to explicitly increase RFP brightness unlike directed evolution and rational design. Since it has previously been successfully used to enhance packing of protein cores[64], this may provide a novel approach to enhance packing of RFP chromophores and therefore their quantum yields to engineer brighter variants, as we will demonstrate in later chapters.

As evidenced in this section, various approaches with different advantages and limitations can be used to engineer RFPs with desired characteristics (**Table 1.2**). However, as shown throughout this Chapter, it remains difficult to engineer RFPs that possess all desired properties combined. As a result, protein engineers focus on improving what is most important for their end applications of interest. Notably, for the majority of these applications, one of the most important properties remains RFP brightness. In this thesis, we will therefore focus our efforts on investigating new rational design methodologies to create bright RFPs.

Table 1.2. Advantages and limitations of different protein engineering approaches.

Approach	Advantages	Limitations
Directed Evolution	• No required knowledge of protein structure or function • Identifies beneficial mutations across the protein structure whose functional effects would be impossible to predict a priori	• Large libraries (> 10^3) require that desired properties be amenable to high-throughput screening • Expensive (time, resources, efforts)
Rational Design	• Fewer mutants to test (< 10^2)	• Requires in-depth knowledge of protein structure and function • Epistasis makes it challenging to accurately predict the functional effect of mutation combinations
CPD	• Can computationally search through large sequence spaces (> 10^5 sequences) to identify mutation combinations that would have been difficult to predict rationally or to obtain through random mutagenesis • Predicts stability of mutants	• Requires large computational resources • Does not always find the global minimum energy conformation • No direct relationship between computed energy and protein function

1.5. Brief Description of Projects

As we discussed in Chapter 1, RFPs generally display low brightness relative to other widely used green and yellow FPs and their brightness tends to decrease the more the emission wavelength is increased. Therefore, the development of methods that increase RFP brightness is desired. One way to go about engineering brighter RFPs is through the rigidification of the chromophore phenolate moiety to limit twisting motions of the chromophore that are responsible for releasing its absorbed energy via non-radiative decay, thus favoring its radiative fluorescence pathway. Herein, we will investigate different rational design strategies based on rigidifying the chromophore to increase brightness in

RFPs. Chapters 2 and 3 will interrogate RFP brightness by introducing mutations directly next to the chromophore, whereas Chapter 4 will focus on distal mutations (**Figure 1.7**). Finally, we will discuss the advantages and limitations of each approach and their contribution to the field. Please note that Chapter 3 will be presented in form of a manuscript to be sent for review to the Journal of the American Chemical Society, and the "materials and methods" sections for Chapters 2 and 4 can be found at the end of this thesis.

In Chapter 2, we hypothesize that the "triple-decker" of aromatic residues engineered by Pandelieva *et al.* (2016)[60] onto the mRojoA scaffold comprising aromatic side chains sandwiching the chromophore phenolate moiety could be improved since the aromatic side chain at position 63 in their construct is oriented edge-to-face with the chromophore. We propose that the quantum yield can therefore be increased by creating a parallel π-stacked "Superdecker" motif comprising R163, H62, the chromophore phenolate moiety and Y197. We used CPD and molecular dynamics simulations to identify stable mutants that contain this motif. Our best variants showed up to 1.3-fold enhanced brightness relative to their parent protein and displayed pH-sensitivity in the alkaline range, potentially making them useful for studying the mitochondrial matrix. We will provide insights on the presence of the *Superdecker* motif in these protein folds and will discuss how the motif can be further engineered to create red-shifted RFPs.

In Chapter 3, we propose an alternate approach to further increase RFP brightness by enhancing chromophore phenolate packing using aliphatic residues instead of aromatic ones and emphasize their ability to create a compact environment. Using CPD, we were able to identify a mutant with 11-fold enhanced brightness from which directed evolution yielded the brightest Discosoma-derived monomeric RFP with an emission maximum

above 600 nm reported to date. We will highlight the benefit of combining CPD and directed evolution to identify improved variants from novel starting points.

In Chapter 4, we test a hypothesis previously developed in our lab whereby RFP brightness can be increased by introducing mutations away from the chromophore environment based on NMR data that revealed a correlation between residue rigidity and RFP quantum yield[65]. We aimed to do so using combinatorial mutagenesis at two different surface patches of three residue positions each on the mCherry template for which the correlation is observed, and whose side chain rotamers are thought to have the ability to interact (distance $_{C\alpha\text{-}C\alpha} \leq 8$ Å) . We will also discuss whether these positions are appropriate targets for this approach and why other positions should also be considered.

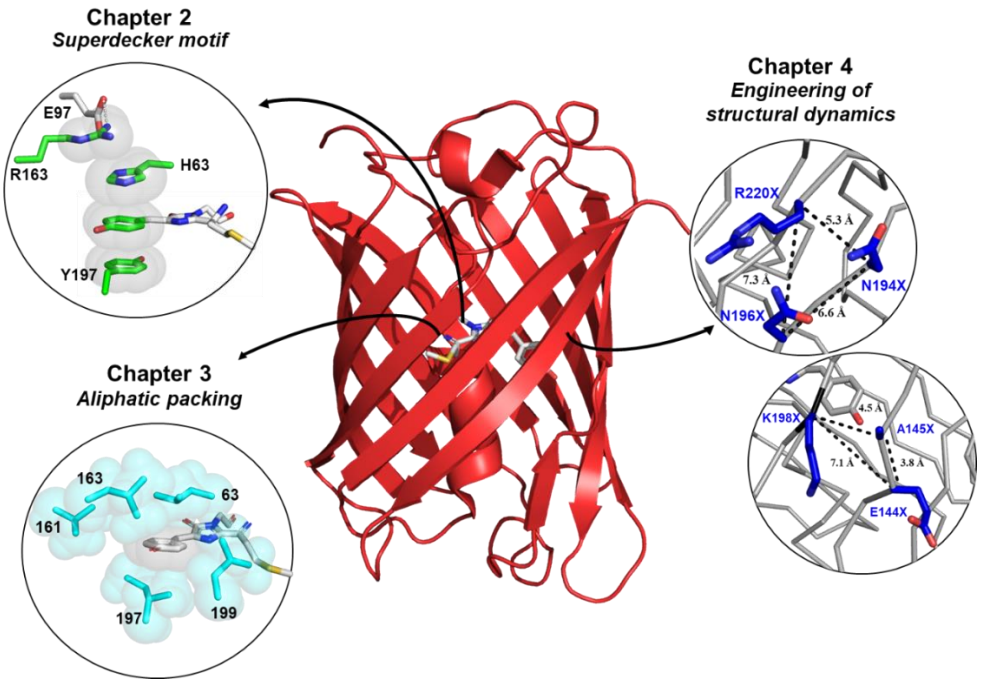


Figure 1.7. Overview of the project objectives. Chapters 2 and 3 focus on engineering RFP brightness by introducing mutations near the chromophore, whereas Chapter 4 relies on mutating distal positions located on the protein surface. The goals of Chapters 2 and 3 are to enhance packing of the chromophore phenolate moiety via a "Superdecker" motif (green) or aliphatic residues (cyan), respectively. The goal of Chapter 4 is to identify beneficial distal mutations where rigidity correlates to the quantum yield using combinatorial mutagenesis of two surface patches (shown in dark blue).

1.6. References

1. Savage, D. C., Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* **1977,** *31*, 107-33.
2. Dantuma, N. P.; Lindsten, K.; Glas, R.; Jellne, M.; Masucci, M. G., Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nat Biotechnol* **2000,** *18* (5), 538-43.
3. Torrance, C. J.; Agrawal, V.; Vogelstein, B.; Kinzler, K. W., Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nat Biotechnol* **2001,** *19* (10), 940-5.
4. SHIMOMURA, O.; JOHNSON, F. H.; SAIGA, Y., Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J Cell Comp Physiol* **1962,** *59*, 223-39.
5. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green fluorescent protein as a marker for gene expression. *Science* **1994,** *263* (5148), 802-5.
6. Heim, R.; Tsien, R. Y., Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol* **1996,** *6* (2), 178-82.
7. Awad, A. I.; Awad, I. A., Chemistry of life: green fluorescent protein and the 2008 Nobel Prize in Chemistry. *Neurosurgery* **2008,** *63* (6), 11.
8. Weiss, P. S., 2008 Nobel Prize in Chemistry: green fluorescent protein, its variants and implications. *ACS Nano* **2008,** *2* (10), 1977.
9. Schnitzler, C. E.; Keenan, R. J.; McCord, R.; Matysik, A.; Christianson, L. M.; Haddock, S. H., Spectral diversity of fluorescent proteins from the anthozoan Corynactis californica. *Mar Biotechnol (NY)* **2008,** *10* (3), 328-42.
10. Mishin, A. S.; Subach, F. V.; Yampolsky, I. V.; King, W.; Lukyanov, K. A.; Verkhusha, V. V., The first mutant of the Aequorea victoria green fluorescent protein that forms a red chromophore. *Biochemistry* **2008,** *47* (16), 4666-73.
11. Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A., Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* **1999,** *17* (10), 969-73.
12. Wiedenmann, J.; Schenk, A.; Röcker, C.; Girod, A.; Spindler, K. D.; Nienhaus, G. U., A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria). *Proc Natl Acad Sci U S A* **2002,** *99* (18), 11646-51.
13. Merzlyak, E. M.; Goedhart, J.; Shcherbo, D.; Bulina, M. E.; Shcheglov, A. S.; Fradkov, A. F.; Gaintzeva, A.; Lukyanov, K. A.; Lukyanov, S.; Gadella, T. W.; Chudakov, D. M., Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods* **2007,** *4* (7), 555-7.
14. Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **2002,** *99* (12), 7877-82.
15. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **2004,** *22* (12), 1567-72.
16. Wang, L.; Jackson, W. C.; Steinbach, P. A.; Tsien, R. Y., Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A* **2004,** *101* (48), 16745-9.
17. Shcherbo, D.; Merzlyak, E. M.; Chepurnykh, T. V.; Fradkov, A. F.; Ermakova, G. V.; Solovieva, E. A.; Lukyanov, K. A.; Bogdanova, E. A.; Zaraisky, A. G.; Lukyanov, S.; Chudakov, D. M., Bright far-red fluorescent protein for whole-body imaging. *Nat Methods* **2007,** *4* (9), 741-6.

18. Kredel, S.; Oswald, F.; Nienhaus, K.; Deuschle, K.; Röcker, C.; Wolff, M.; Heilker, R.; Nienhaus, G. U.; Wiedenmann, J., mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PLoS One* **2009,** *4* (2), e4391.
19. Shcherbo, D.; Murphy, C. S.; Ermakova, G. V.; Solovieva, E. A.; Chepurnykh, T. V.; Shcheglov, A. S.; Verkhusha, V. V.; Pletnev, V. Z.; Hazelwood, K. L.; Roche, P. M.; Lukyanov, S.; Zaraisky, A. G.; Davidson, M. W.; Chudakov, D. M., Far-red fluorescent tags for protein imaging in living tissues. *Biochem J* **2009,** *418* (3), 567-74.
20. Lam, A. J.; St-Pierre, F.; Gong, Y.; Marshall, J. D.; Cranfill, P. J.; Baird, M. A.; McKeown, M. R.; Wiedenmann, J.; Davidson, M. W.; Schnitzer, M. J.; Tsien, R. Y.; Lin, M. Z., Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat Methods* **2012,** *9* (10), 1005-12.
21. Bajar, B. T.; Wang, E. S.; Lam, A. J.; Kim, B. B.; Jacobs, C. L.; Howe, E. S.; Davidson, M. W.; Lin, M. Z.; Chu, J., Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci Rep* **2016,** *6*, 20889.
22. Li, Z.; Zhang, Z.; Bi, L.; Cui, Z.; Deng, J.; Wang, D.; Zhang, X. E., Mutagenesis of mNeptune Red-Shifts Emission Spectrum to 681-685 nm. *PLoS One* **2016,** *11* (4), e0148749.
23. Matela, G.; Gao, P.; Guigas, G.; Eckert, A. F.; Nienhaus, K.; Nienhaus, G. U., A far-red emitting fluorescent marker protein, mGarnet2, for microscopy and STED nanoscopy. *Chem Commun (Camb)* **2017,** *53* (5), 979-982.
24. Bindels, D. S.; Haarbosch, L.; van Weeren, L.; Postma, M.; Wiese, K. E.; Mastop, M.; Aumonier, S.; Gotthard, G.; Royant, A.; Hink, M. A.; Gadella, T. W., mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **2017,** *14* (1), 53-56.
25. Bischof, H.; Burgstaller, S.; Waldeck-Weiermair, M.; Rauter, T.; Schinagl, M.; Ramadani-Muja, J.; Graier, W. F.; Malli, R., Live-Cell Imaging of Physiologically Relevant Metal Ions Using Genetically Encoded FRET-Based Probes. *Cells* **2019,** *8* (5).
26. Ransom, E. M.; Ellermeier, C. D.; Weiss, D. S., Use of mCherry Red fluorescent protein for studies of protein localization and gene expression in Clostridium difficile. *Appl Environ Microbiol* **2015,** *81* (5), 1652-60.
27. Nasu, Y.; Shen, Y.; Kramer, L.; Campbell, R. E., Structure- and mechanism-guided design of single fluorescent protein-based biosensors. *Nat Chem Biol* **2021,** *17* (5), 509-518.
28. Qian, Y.; Cosio, D. M. O.; Piatkevich, K. D.; Aufmkolk, S.; Su, W. C.; Celiker, O. T.; Schohl, A.; Murdock, M. H.; Aggarwal, A.; Chang, Y. F.; Wiseman, P. W.; Ruthazer, E. S.; Boyden, E. S.; Campbell, R. E., Improved genetically encoded near-infrared fluorescent calcium ion indicators for in vivo imaging. *PLoS Biol* **2020,** *18* (11), e3000965.
29. Shen, Y.; Wu, S. Y.; Rancic, V.; Aggarwal, A.; Qian, Y.; Miyashita, S. I.; Ballanyi, K.; Campbell, R. E.; Dong, M., Genetically encoded fluorescent indicators for imaging intracellular potassium ion concentration. *Commun Biol* **2019,** *2*, 18.
30. Clement, G. T.; Hynynen, K., Ultrasound phase-contrast transmission imaging of localized thermal variation and the identification of fat/tissue boundaries. *Phys Med Biol* **2005,** *50* (7), 1585-600.
31. Ash, C.; Dubec, M.; Donne, K.; Bashford, T., Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers Med Sci* **2017,** *32* (8), 1909-1918.
32. Yang, M.; Jiang, P.; Yamamoto, N.; Li, L.; Geller, J.; Moossa, A. R.; Hoffman, R. M., Real-time whole-body imaging of an orthotopic metastatic prostate cancer model expressing red fluorescent protein. *Prostate* **2005,** *62* (4), 374-9.

33. Winnard, P. T.; Kluth, J. B.; Raman, V., Noninvasive optical tracking of red fluorescent protein-expressing cancer cells in a model of metastatic breast cancer. *Neoplasia* **2006,** *8* (10), 796-806.
34. Strack, R. L.; Strongin, D. E.; Mets, L.; Glick, B. S.; Keenan, R. J., Chromophore formation in DsRed occurs by a branched pathway. *J Am Chem Soc* **2010,** *132* (24), 8496-505.
35. Moore, M. M.; Oteng-Pabi, S. K.; Pandelieva, A. T.; Mayo, S. L.; Chica, R. A., Recovery of red fluorescent protein chromophore maturation deficiency through rational design. *PLoS One* **2012,** *7* (12), e52463.
36. Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J., Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996,** *273* (5280), 1392-5.
37. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-47.
38. Sniegowski, J. A.; Lappe, J. W.; Patel, H. N.; Huffman, H. A.; Wachter, R. M., Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *J Biol Chem* **2005,** *280* (28), 26248-55.
39. Barondeau, D. P.; Putnam, C. D.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D., Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc Natl Acad Sci U S A* **2003,** *100* (21), 12111-6.
40. Regmi, C. K.; Bhandari, Y. R.; Gerstman, B. S.; Chapagain, P. P., Exploring the diffusion of molecular oxygen in the red fluorescent protein mCherry using explicit oxygen molecular dynamics simulations. *J Phys Chem B* **2013,** *117* (8), 2247-53.
41. Terskikh, A. V.; Fradkov, A. F.; Zaraisky, A. G.; Kajava, A. V.; Angres, B., Analysis of DsRed Mutants. Space around the fluorophore accelerates fluorescence development. *J Biol Chem* **2002,** *277* (10), 7633-6.
42. Wenzel, T.; Dana, C. A. Quantum Yield of Fluorescence. https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Book%3A_Molecular_and_Atomic_Spectroscopy_(Wenzel)/3%3A_Molecular_Luminescence/3.5._Quantum_Yield_of_Fluorescence_((varphi_ce_F)) (accessed May 31, 2021).
43. Wong, D. Fluorescence and Phosphorescence. https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Electronic_Spectroscopy/Fluorescence_and_Phosphorescence (accessed June 17, 2021).
44. Haebig, J. E., Vibrational effects in radiationless transitions in aromatic molecules. *J Phys Chem* **1967,** *71* (13), 4203-4209.
45. Garcia-Parajo, M. F.; Segers-Nolten, G. M.; Veerman, J. A.; Greve, J.; van Hulst, N. F., Real-time light-driven dynamics of the fluorescence emission in single green fluorescent protein molecules. *Proc Natl Acad Sci U S A* **2000,** *97* (13), 7237-42.
46. Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A., Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* **2010,** *90* (3), 1103-63.
47. Wu, F.; Van Rijn, E.; Van Schie, B. G.; Keymer, J. E.; Dekker, C., Multi-color imaging of the bacterial nucleoid and division proteins with blue, orange, and near-infrared fluorescent proteins. *Front Microbiol* **2015,** *6*, 607.
48. Thorn, K., Genetically encoded fluorescent tags. *Mol Biol Cell* **2017,** *28* (7), 848-857.
49. Chu, J.; Haynes, R. D.; Corbel, S. Y.; Li, P.; González-González, E.; Burg, J. S.; Ataie, N. J.; Lam, A. J.; Cranfill, P. J.; Baird, M. A.; Davidson, M. W.; Ng, H. L.; Garcia, K. C.; Contag, C. H.; Shen, K.; Blau, H. M.; Lin, M. Z., Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nat Methods* **2014,** *11* (5), 572-8.

50. Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L., Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc Natl Acad Sci U S A* **2010,** *107* (47), 20257-62.
51. Suzuki, T.; Arai, S.; Takeuchi, M.; Sakurai, C.; Ebana, H.; Higashi, T.; Hashimoto, H.; Hatsuzawa, K.; Wada, I., Development of cysteine-free fluorescent proteins for the oxidative environment. *PLoS One* **2012,** *7* (5), e37551.
52. Drobizhev, M.; Molina, R. S.; Callis, P. R.; Scott, J. N.; Lambert, G. G.; Salih, A.; Shaner, N. C.; Hughes, T. E., Local Electric Field Controls Fluorescence Quantum Yield of Red and Far-Red Fluorescent Proteins. *Front Mol Biosci* **2021,** *8*, 633217.
53. Englman, R.; Jortner, J., The energy gap law for radiationless transitions in large molecules. *Mol Phys* **1970,** *18* (2), 145-164.
54. Wei, Y.-C.; Wang, S. F.; Hu, Y.; Liang-Sheng, L.; Chen, D.-G.; Chang, K.-H.; Wang, C.-W.; Liu, S.-H.; Chan, W.-H.; Liao, J.-L.; Hung, W.-Y.; Wang, T.-H.; Chen, P.-T.; Hsu, H.-F.; Chi, Y.; Chou, P.-T., Overcoming the energy gap law in near-infrared OLEDs by exciton–vibration decoupling. *Nat Photonics* **2020,** *14*, 570–577.
55. Baffour-Awuah, N. Y. A.; Zimmer, M., Hula-twisting in green fluorescent protein. *Chem Phys* **2004,** *303* (1–2), 7-11.
56. Olsen, S.; Smith, S. C., Radiationless decay of red fluorescent protein chromophore models via twisted intramolecular charge-transfer states. *J Am Chem Soc* **2007,** *129* (7), 2054-65.
57. Chen, K.; Arnold, F. H., Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc Natl Acad Sci U S A* **1993,** *90* (12), 5618-22.
58. Scott, J. K.; Smith, G. P., Searching for peptide ligands with an epitope library. *Science* **1990,** *249* (4967), 386-90.
59. Winter, G.; Griffiths, A. D.; Hawkins, R. E.; Hoogenboom, H. R., Making antibodies by phage display technology. *Annu Rev Immunol* **1994,** *12*, 433-55.
60. Pandelieva, A. T.; Baran, M. J.; Calderini, G. F.; McCann, J. L.; Tremblay, V.; Sarvan, S.; Davey, J. A.; Couture, J. F.; Chica, R. A., Brighter Red Fluorescent Proteins by Rational Design of Triple-Decker Motif. *ACS Chem Biol* **2016,** *11* (2), 508-17.
61. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.
62. Wannier, T. M.; Moore, M. M.; Mou, Y.; Mayo, S. L., Computational Design of the β-Sheet Surface of a Red Fluorescent Protein Allows Control of Protein Oligomerization. *PLoS One* **2015,** *10* (6), e0130582.
63. Wannier, T. M.; Gillespie, S. K.; Hutchins, N.; McIsaac, R. S.; Wu, S. Y.; Shen, Y.; Campbell, R. E.; Brown, K. S.; Mayo, S. L., Monomerization of far-red fluorescent proteins. *Proc Natl Acad Sci U S A* **2018,** *115* (48), E11294-E11301.
64. Allen, B. D.; Nisthal, A.; Mayo, S. L., Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proc Natl Acad Sci U S A* **2010,** *107* (46), 19838-43.
65. Damry, A. M. From protein sequence to motion to function: Towards the rational design of functional protein dynamics. Doctoral dissertation, Université d'Ottawa/University of Ottawa, 2019.

Chapter 2: Engineering a *Superdecker* motif of parallel π-stacked aromatic residues

Sandrine Legault[1], Niayesh Zarifi[1], Ralph McAnelly[2], Michael C. Thompson[2,3], James S. Fraser[3] & Roberto A. Chica[1]

[1]Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie-Curie, Ottawa, Ontario, Canada, K1N 6N5

[2]Department of Chemistry and Biochemistry, University of California Merced, 5200 North Lake Rd. Merced, CA 95343

[3]Department of Bioengineering and Therapeutic Sciences California Institute of Quantitative Biosciences (QB3), 600 16th Street, San Francisco, CA 94143

2.1. Statement of Contribution

Crystallography of mRojo-SD1 was performed by our collaborators Dr. Michael C. Thompson and Dr. James S. Fraser, and crystallography of our *Superdecker* mutants is currently underway by Ralph McAnelly and Dr. Michael C. Thompson. Ensemble refinement of mRojo-SD1 and molecular dynamics simulations of designed *Superdecker* mutants were performed by Niayesh Zarifi. All other computational design calculations were performed by Dr. Roberto A. Chica. Experimental design was done by Dr Roberto A. Chica and Sandrine Legault, and all experimental laboratory work was performed by Sandrine Legault.

Acknowledgements. S.L. is the recipient of an Ontario Graduate Scholarship and a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada. We thank Noreen Ahmed for help with protein purification, and John P. Pezacki for use of the fast protein liquid chromatography system and Molecular Devices plate reader.

2.2. Introduction

Red fluorescent proteins (RFPs) are genetically-encoded fluorophores that have found widespread application in chemical and biological research due to their longer emission wavelengths that allow deeper tissue imaging[1] and multicolor experiments[2, 3]. For most applications, bright RFPs are desirable. However, most RFPs display lower brightness relative to other widely used green or yellow fluorescent proteins, emphasizing the need to develop methods to increase RFP brightness. RFP brightness is typically improved by directed evolution involving multiple rounds of random mutagenesis followed by high-throughput screening of large mutant libraries ($\geq 10^3$ sequences per round)[4, 5]. Although directed evolution has successfully yielded many RFPs displaying enhanced brightness, a rational methodology to increase brightness would be useful to generate novel bright variants without the need for screening large mutant libraries, while allowing us to test our understanding of the structural basis of high brightness in these proteins

As discussed in Chapter 1, brightness of fluorescent proteins is defined as the product of their molar extinction coefficient and quantum yield, which are the ability of their chromophore to absorb light and its efficiency to convert the absorbed light into emitted light, respectively. Although increasing either of these two properties will proportionally increase brightness, it is not well understood how changes to the RFP structure can beneficially impact its extinction coefficient, complicating the prediction of beneficial mutations by rational design. On the other hand, it is known that the quantum yield of fluorophores is directly related to conformational flexibility[6-8], as torsional motions can dissipate the absorbed energy as heat instead of as photons. This process is known as twisting intramolecular charge transfer[5] (TICT). In the case of fluorescent proteins, it has

been shown that twisting of the chromophore phenolate moiety via torsions of the CB2-CG2 bond causes non-radiative decay[8, 9]. Therefore, it should be possible to enhance RFP brightness by designing mutations to restrict the conformational flexibility of the phenolate moiety, resulting in higher quantum yield.

Here, we propose that introduction of a motif comprising a parallel π-stacking of the chromophore phenolate moiety between aromatic side chains has the ability to rigidify the chromophore and increase the brightness of a previously engineered RFP, mRojo-VHSV[6]. This RFP contains aromatic residues π-stacked with the chromophore in an edge-to-face arrangement that we hypothesize is not optimal for preventing phenol ring twisting. In order to introduce a face-to-face orientation, we aim to engineer a motif comprising the chromophore phenolate moiety π-stacked between a tyrosine and a histidine that is oriented in a parallel orientation via a cation-π interaction with the guanidium group of an arginine side chain, which we named *Superdecker* (**Figure 2.1c**). Using computational protein design on a fast- and efficiently-maturing mRojo-VHSV variant we were able to identify sequences in which our motif was predicted to be stable. The best mutants identified, mSD1.1 and mSD1.5, display up to 1.7-fold increases in quantum yield resulting in a modest 1.3-fold enhanced brightness at pH 9.0. They also display high pK_a values of 8.1 and 8.3, respectively. Crystallography of mSD1.5 is currently underway to confirm the presence of the Superdecker motif, which we found by MD simulations to be relatively more stable in mSD1.5 than the other designed variants.

2.3. Results

Previous work, rationale and design hypothesis

Previous work in our lab demonstrated that the quantum yield (Φ_F) of the dim red-shifted mCherry variant mRojoA (Φ_F = 0.024) can be increased by sandwiching the chomophore phenolate moiety between the side chain of Y197, which is already present in mRojoA, and of another aromatic residue at position 63 above the chromophore (P63Y, P63H or P63F), creating a *triple-decker* motif[6]. The brightest triple-decker variant, mRojo-VHSV, which is named based on its mutations relative to mRojoA (V16, H63, S143, V163), displays more than 2-fold increased brightness and 1.7-fold increased quantum yield (Φ_F = 0.040). However, similar to the other *triple-decker* mutants, the side chain of the aromatic residue above the chromophore, in this case H63, points in the same direction as the chromophore phenolate but is not stacked in a parallel orientation relative to the chromophore unlike Y197 (**Figure 2.1a**).

We hypothesized that the edge-to-face orientation of H63 in mRojo-VHSV results in sub-optimal packing and allows twisting of the phenol ring of the chromophore resulting in non-radiative decay[5]. Therefore, we proposed that quantum yield and brightness could be further increased by engineering a parallel π-stacked motif to enhance packing of the chromophore phenolate moiety. A similar motif is observed in the Copepod GFP, which contains a parallel histidine that is π-stacked above the green chromophore[7]. In this protein, the histidine is oriented via cation-π interactions with an arginine, which itself is stabilized via hydrogen-bonding with a glutamate (**Figure 2.1b**).

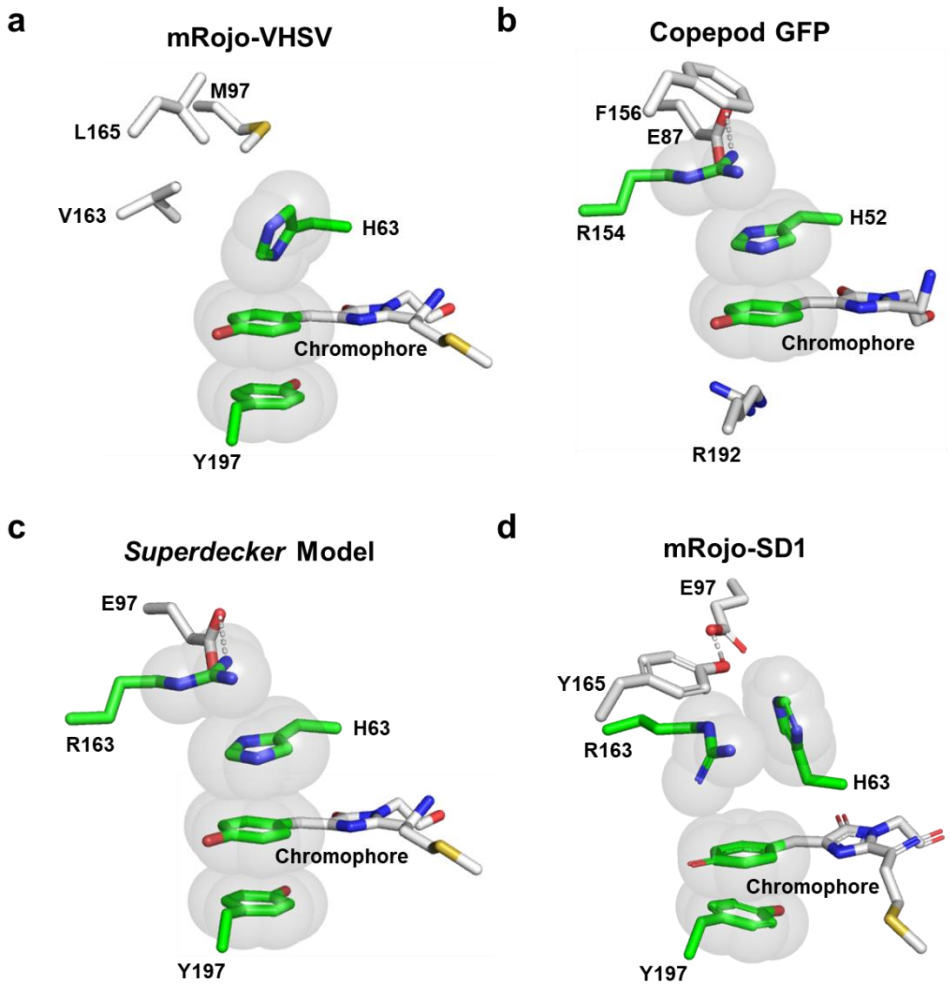


Figure 2.1. Rational design of a *Superdecker* motif. (a) mRojo-VHSV (adapted from mRojo-THSL PDB ID: 5H87[6], in which T16V and L163V were introduced using the mutation wizard in PyMOL), (b) Copepod GFP (PDB ID: 2DD7[7]), (c) hypothetical *Superdecker* model, (d) mRojo-SD1 (not published). The residues and the chromophore phenolate moiety comprising the Superdecker motif are shown as green sticks, the hydrogen bonds as white dotted lines, and the packing interactions are represented by gray spheres.

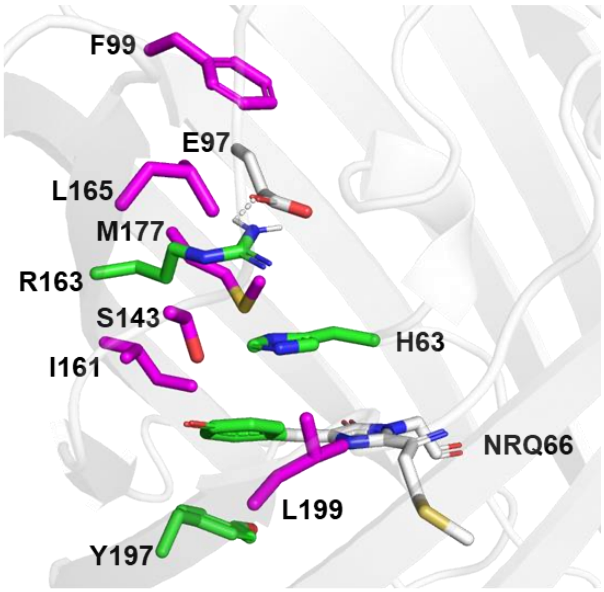
To create an RFP containing a triple-decker motif in which the aromatic plane of the H63 side chain is parallel to that of the chromophore phenolate moiety, we first introduced homologous mutations to the Copepod GFP (M97E and V163R) into mRojo4.3, a fast- and efficient-maturing variant of mRojo-VHSV that was previously evolved in our

lab. We named this motif a *Superdecker* (**Figure 2.1c).** However, crystallographic data showed that the parallel motif was not present in our construct which we named mRojo-SD1, where SD1 refers to ***S****uper****D****ecker* mutant **1** (**Figure 2.1d**). This suggested that residues surrounding the motif should also be considered in the design to allow the desired geometry as some first shell residue conformers may participate in steric hindrance of residues of the motif or may disturb necessary electrostatic interactions, for example E97 was shown to make a hydrogen bond with Y165 instead of R163 in mRojo-SD1, thereby not allowing them to adopt a parallel arrangement.

Computational protein design

In order to create the *Superdecker* motif, we turned to computational protein design (CPD). We first generated a backbone ensemble of mRojo-SD1 using molecular dynamics restrained by the diffraction data, also known as ensemble refinement (see **5.1. Materials and Methods**). This procedure produced an ensemble of 63 templates approximating the conformational flexibility of the mRojo-SD1 scaffold. Using this ensemble, we designed six residues (99, 143, 161, 165, 177 and 199) in order to stabilize the Superdecker motif comprising Y197, the phenolate moiety of the chromophore, H63, and R163 (**Figure 2.2a**).

a



b

		Positions										
Seq	Energy	63	66	97	99	143	161	163	165	177	197	199
WT		H	NRQ	E	F	S	I	R	Y	V	Y	M
mSD1.0	-2540.25581	-	*	-	-	-	M	-	L	I	-	L
mSD1.1	-2535.51895	-	*	-	-	-	-	-	L	M	-	L
mSD1.2	-2530.48440	-	*	-	-	-	-	-	F	M	-	L
mSD1.3	-2528.18771	-	*	-	-	-	-	-	M	M	-	L
mSD1.4	-2526.58491	-	*	-	-	D	M	-	F	-	-	L
mSD1.5	-2526.15859	-	*	-	-	-	L	-	L	M	-	L
mSD1.6	-2523.17043	-	*	-	M	-	-	-	L	-	-	L
mSD1.7	-2522.82967	-	*	-	-	D	L	-	M	-	-	L

Figure 2.2. Computational design of *Superdecker* mutant mSD1.1. (a) Computational design model of mutant mSD1_1. Designed residues are shown as magenta sticks and the hydrogen bond between E97 and R163 is shown as a white dotted line. NRQ indicates the RFP chromophore. (b) CPD results. Asterisks, letters and dashes indicate the chromophore, mutations, or wild-type residues, respectively. The sequences shown correspond to the top 8 sequences, ranked according to their Boltzmann-weighted average energy (T = 300K) across the backbone template ensemble, that were predicted to adopt the *Superdecker* motif as their most stable conformation. The designed sequence of mSD1.1 is highlighted in gray.

From the top ranked 100 designed sequences, 8 were predicted to adopt the *Superdecker* in their most stable predicted structure. These sequences were named based on their ranking from the CPD calculation. For example, mSD1.0 refers to the mRojo-SD1 mutant with the lowest calculated potential energy value, mSD1.1 refers the second ranked mutant, etc. These sequences contain between 3 and 4 mutations relative to mRojo-SD1: a

leucine at position 199, whose branched side chain (isobutyl moiety) may contribute to enhanced packing, and various combinations of residues at the other five design positions, including hydrophobic residues at position 165 which should not interfere with the hydrogen bonding between E97 and R163 since they cannot act as competing hydrogen bond donors or acceptors. (**Figure 2.2b**). In addition, the R125H surface mutation found in mRojoA and derivatives was also reverted to enhance solubility and favor the monomeric state of our mutants[8] (**Supplementary Figure S2.1**). These designed sequences were selected for experimental characterization.

Experimental screening and spectroscopic characterization

Only three of the eight computationally designed mutants displayed fluorescence distinguishable from the background noise (**Supplementary Figure S2.2**). Mutant mSD1.0 exhibited enhanced fluorescence intensity relative to the parent mRojo-SD1 with a similar emission maximum at approximately 613 nm. Interestingly, the emission spectra of mutants mSD1.1 and mSD1.5 displayed emission peaks that were blue-shifted by 17 and 7 nm, respectively. Presence of this blue-shifted emission suggests a change in the electronic environment of the chromophore, which could potentially be caused by the presence of the *Superdecker*. In support of this assumption, Khrenova *et al.* (2020) predicted by quantum mechanical calculations that a 0.10 eV (approximately 30-nm) hypsochromic shift in emission wavelength would result from a His-chromophore motif in which the aromatic rings adopt parallel orientations relative to each other[9]. All other variants displayed poor fluorescence, potentially resulting from hindered red chromophore

maturation. Therefore, to facilitate experimental manipulation, the three mutants with enhanced fluorescence intensity were chosen for further characterization.

All three mutants displayed efficient red chromophore formation, similar to their parent protein mRojo-SD1. This is evidenced by the lack of green chromophore peaks (λ_{abs} = 390 nm and 510 nm)[10] in all absorption spectra (**Figure 2.3**).

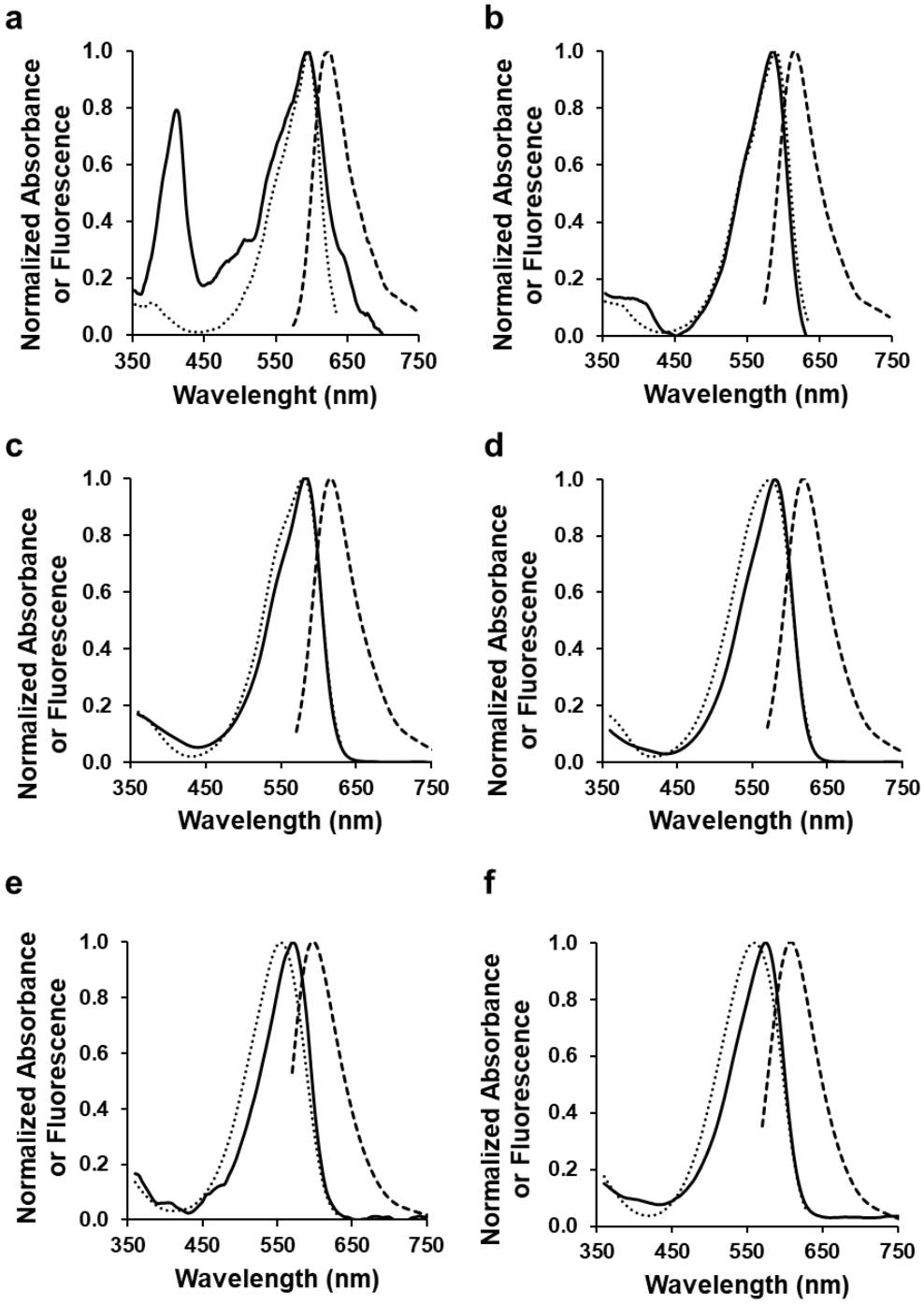


Figure 2.3. Fluorescence and absorption spectra of *Superdecker* RFPs. Absorption (solid lines), excitation (dotted lines) and emission (dashed lines) spectra of purified (a) mRojo-VHSV, (b) Mutant4.3, (c) mRojo-SD1, (d) mSD1.0, (e) mSD1.1 and (f) mSD1.5 in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4). Spectra were measured at room temperature using an Infinite M1000 (Tecan) or Spectramax i3 (Molecular Devices) plate reader.

At physiological pH, the brightest variant mSD1.1 displayed an absolute increase in quantum yield of 0.012 (120%) relative to the parent mRojo-SD1 (**Table 2.1**) whereas the quantum yields of mSD1.0 and mSD1.5 were considerably lower. However, the lower extinction coefficient of mSD1.1 did not result in enhanced overall brightness. Interestingly, these three designed mutants all displayed altered pH profiles relative to mRojo-VHSV and mRojo-SD1 (**Supplementary Figure S2.3**). All mutants displayed a single pK_a value between 8.1 and 8.7, which are higher than what is usually observed in RFPs (reported pK_a's for most RFPs are in the 4.0–6.5 range)[11]. Because of the pH sensitivity in the alkaline range identified in the *Superdecker* variants, spectral characterization was also performed at pH 9.0. At this pH, mSD1.1 and mSD1.5 exhibited 1.7- and 1.4-fold increases in quantum yield relative to mRojo-SD1, respectively, resulting in approximately 1.3-fold enhanced brightness.

Table 2.1. Spectral properties of *Superdecker* RFPs.

Protein	λ_{abs} (nm)	λ_{ex} (nm)	λ_{em} (nm)	Quantum Yield	Extinction Coefficient ($mM^{-1}cm^{-1}$)	Brightness ($mM^{-1}cm^{-1}$)	pKa	Mutations Relative to mRojo-VHSV									
								97	125	150	161	163	165	177	194	199	217
mRojo-VHSV*	-	592 ± 1	622 ± 2	0.040 ± 0.004	117 ± 13	4.7	4.3, 9.5	M	H	M	I	V	L	V	N	L	C
Mutant4.3	586 ± 1	591 ± 1	617 ± 1	0.047 ± 0.001	73 ± 6	3.4	-	-	-	L	-	-	Y	-	Y	-	S
mRojo-SD1	585 ± 1 (565)	585 ± 1 (570)	612 ± 3 (610)	0.073 ± 0.002 (0.083)	93 ± 4	6.8 (7.7)	4.5, 8.6	E	-	L	-	R	Y	-	Y	M	S
mSD1.0	583 (550)	575 (550)	613 (605)	0.034 (0.076)	84	2.9 (6.4)	8.7	E	R	L	M	R	-	I	Y	-	S
mSD1.1	565 (545)	555 (545)	595 (595)	0.085 (0.14)	66	5.6 (9.2)	8.1	E	R	L	-	R	-	M	Y	-	S
mSD1.5	580 (545)	560 (545)	605 (595)	0.050 (0.12)	83	4.2 (10)	8.2	E	R	L	L	R	-	M	Y	-	S

λ_{abs}, λ_{ex} and λ_{em} correspond to the absorption, excitation, and emission maxima, respectively.
Values in parentheses are at pH 9.0 (0.02 M sodium phosphate buffer).
All other values are at pH 7.4 in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate).
Brightness is the product of the quantum yield and the extinction coefficient.
The pK_a is the pH value at which 50% of the maximal fluorescence is obtained. mRojo-VHSV and mRojo-SD1 have two pK_a values (**Supplementary Figure S2.3**).
n = 2 independent experiments for Mutant4.3 and mRojo-SD1. n = 1 independent experiment for all other variants. Each independent experiment consisted of triplicate measurements performed on a single protein batch.
*Data from Pandelieva *et al.* (2016)[6].

Structural analysis

To evaluate the relative stability of the *Superdecker* motif in the designs of mSD1.0, mSD1.1 and mSD1.5 variants, we performed molecular dynamics (MD) simulations over a period of 200 ns on each computational model. MD is a computational tool that simulates the movement of atoms in real-time to evaluate the evolution of molecular interactions within the protein over a chosen timescale. We chose a period in the short nanosecond range as this timescale allows us to visualize side chain motions[12]. We expect that a stable *Superdecker* motif would preserve the parallel orientation of side chain rotamers during this simulation time, whereas an unstable one would not. Our results demonstrated that the *Superdecker* motif was preserved in mSD1.5 (**Figure 2.4**), whereas the residues involved in the motif in mSD1.0 and mSD1.1 shifted to conformations similar to those seen in the crystal structure of mRojo-SD1 (**Figure 2.1d**), which are not parallel.

In addition to visual inspection of the MD trajectory, we performed a more quantitative analysis of the parallel π-stacking between H63 and the phenolate moiety of the chromophore by plotting the distance between the ring centroids, as well as the angle between the aromatic planes, in function of the simulation time course (**Supplementary Figure S2.4a**). From the simulations, we observed that the H63 and the chromophore phenolate in mSD1.5 were in the desired geometry, where their centroid distance varied around 4.0 Å, which is close to the ideal distance for π-stacking interactions (3.5 Å)[13], and the aromatic rings maintained their parallel arrangement with an angle between the aromatic planes that did not exceed 25 degrees (**Supplementary Figure S2.4b**). In contrast, the angle between aromatic planes in mSD1.0 and mSD1.1 fluctuated between 20

and 90 degrees over the course of the simulation, suggesting that the parallel *Superdecker* motif is relatively less stable in these variants.

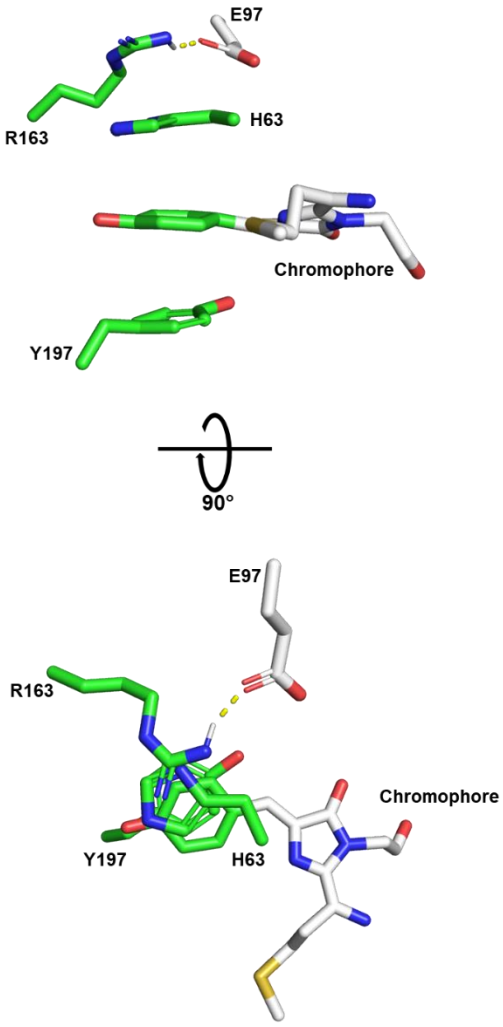


Figure 2.4. Molecular dynamics simulation snapshot of mSD1.5. MD simulation of mSD1.5 were performed using the using the Amber 2018 software (see Chapter 5, Materials and Methods). At the end of the simulation, 1000 snapshots separated by 0.2 ns along the production trajectory were extracted. Snapshots of the design model of mSD1.5 revealed that the parallel stacking of the *Superdecker* motif (shown as green sticks) was retained over the course of the 200 ns simulation. The model structure shown here corresponds to snapshot 625 out of 1000.

The blue-shifted emission, high pKa values, and higher brightness at pH 9.0 observed for the *Superdecker* variants suggest that we have significantly altered the stereoelectronic environment of the chromophore due to the designed mutations. Molecular dynamics simulations suggest that the *Superdecke*r motif is more stable in some mutants than others. This observation, coupled with the observed hypsochromic shift by Khrenova et al. (2020), supports the hypothesis that the motif may be present in some variants. To evaluate the presence of the target *Superdecker* motif, we purified mSD1.0, mSD1.1 and mSD1.5 and sent them to our collaborator Michael C. Thompson (UC Merced) for crystallography, which is currently underway.

2.4. Discussion

In this work, we used CPD as a rational design approach to identify RFP sequences predicted to contain a *Superdecker* motif composed of a parallel side chain stacking of Y197, the chromophore phenolate moiety, H63 and R163, designed to enhance packing of the chromophore and thereby increase RFP brightness. Using this methodology, we were able to identify three efficiently-maturing mutants: mSD1.0, mSD1.1 and mSD1.5. mSD1.0 was dimmer than the parent mRojo-SD1 and displayed similar excitation and emission wavelengths. In contrast, mSD1.1 exhibited an increased quantum yield of 1.2-fold at physiological pH. However, its lower extinction coefficient, which may result from a mixture of protonated and deprotonated chromophores in the RFP population, did not result in improved overall brightness. On the other hand, at pH 9.0, mSD1.1 and mSD1.5 displayed quantum yields up to 1.7-fold higher than mRojo-SD1, which successfully yielded 1.3-fold enhanced brightness. These brighter and pH-sensitive variants also

displayed 15 nm blue-shifted emission maxima, and 20 nm blue-shifted absorption maxima at pH 9.0, relative to mRojo-SD1. Similar hypsochromic shifts were also observed at pH 7.4.

A shift in wavelength is indicative of a change in the electronic environment of the chromophore. For RFP applications, a bathochromic shift is usually desired and we will discuss in the conclusion how our motif can be engineered for this purpose. However, in the case of the work herein, the hypsochromic shifts may suggest the presence of the desired *Superdecker* in which H63 π-stacks with the chromophore phenolate moiety in a face-to-face orientation. This is supported by quantum chemistry simulations results obtained by Khrenova *et al.* (2020) where they developed a quantitative structure-property relationship model to predict spectral tuning in FPs and demonstrated that parallel π-stacking interactions between the DsRed-like RFP chromophore and a histidine, which they substituted as an imidazole, results in a hypsochromic shift in the absorption spectrum[9].

In addition, we used MD simulations on the design structure models of mSD1.0 and mSD1.5 obtained by CPD to evaluate the relative stability of the Superdecker motif in mSD1.0, mSD1.1 and mSD1.5 computational models. From a structural standpoint, the goal of engineering the *Superdecker* was to force H63 in a parallel orientation relative to the chromophore phenolate moiety so as to enhance its packing and restrict twisting motions resulting in TCIT. Therefore, we evaluated both the possibility of interactions between the two rings by verifying that the distance between the ring centroids was ~ 3.5 Å and the parallel arrangement by verifying that the angle value between the aromatic

planes was less than 25°. We observed that the criteria and therefore the motif was retained in mSD1.5 over the course of the simulation, but not in mSD1.0 or mSD1.1.

There is a discrepancy between the results obtained by CPD calculations which predicted the presence of the *Superdecker* in the design sequences including mSD1.0 and what was obtained using MD simulations. This is due to the fact that the advantage of CPD is to identify stable conformations by finding the best combination of rotamers, whereas MD takes into account protein dynamics and calculates the forces on each atom in order to predict their movement in real-time. In other words, the *Superdecker* in mSD1.0 is predicted to be stable (as shown by CPD results) but in real-time the atoms of the *Superdecker* residues may also be subjected to forces such that the structural motif may escape the local minimum of its energy well and so may only be present during a negligible amount of time (as shown in MD results).

Apart from structural insights, an interesting spectral characteristic that was observed in the mutants is their high pK_a values ranging from 8.1 to 8.7 and which confers mSD1.1 and mSD1.5 their significant quantum yield increases and thereby brightness enhancement at pH 9.0. mRojo-VHSV and its derivative mRojo-SD1 both possess two pK_a values, one around 4 which is typically within the range reported for RFPs[11] and represents the ionizable phenol of the chromophore and the second pK_a around 9. The nature of this second pK_a is not well understood but is thought to represent an additional ionisable group on the chromophore, potentially the secondary amine of the imidazolinone which may participate in hydrogen bonding with the carboxylate of the conserved Glu215 side chain in some RFPs[14]. The single high pK_a values in the *Superdecker* mutants seem to indicate that their electrostatic chromophore environments have been affected such that the phenol

moiety is less acidic and thus more resistant to deprotonation. As illustrated in **Figure 1.3** of Chapter 1, the fluorescent RFP chromophore is the deprotonated form which would explain why the Superdecker mutants are dimmer at physiological pH relative to their parent mRojo-SD1 which has a pK_a of 4.5. Nevertheless, the unusually high pK_a of the ionisable chromophore phenol moiety of the *Superdecker* mutants could be useful to monitor processes that occur at pH in the alkaline range.

2.5. Conclusions and Perspectives

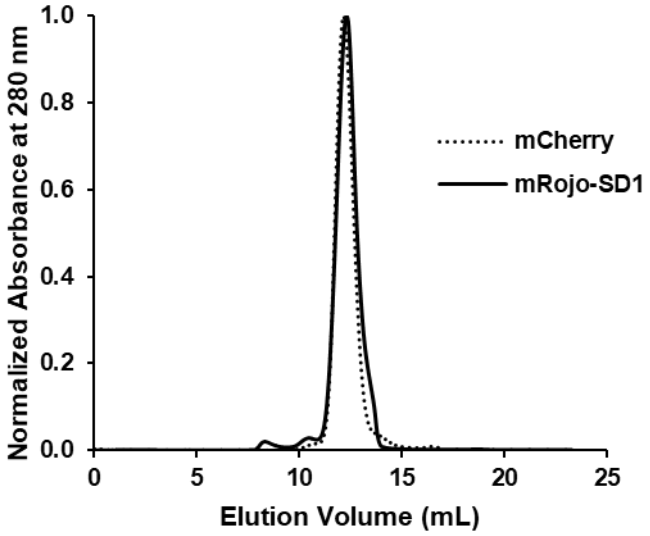
Using CPD, we generated *Superdecker* variants that are brighter than their parent protein at pH 9.0. The pH-sensitivity of our mutants in the alkaline range, not usually observed in RFPs, could potentially render them useful for studying the dynamics of the mitochondrial matrix (pH ~ 8.5)[15].

To validate the presence of our designed parallel structural *Superdecker* motif comprising Y197, the chromophore phenolate moiety, H63 and R163, crystallography of mSD1.0, mSD1.1 and mSD1.5 mutants is currently underway. MD was also used to evaluate the relative stability of the motif in mutants' computational models. The motif was observed for the majority of the simulations for mSD1.5 but not for mSD1.0 or mSD1.1, suggesting that the *Superdecker* motif is more stable in some mutants. This observation, coupled with the observed hypsochromic shift by Khrenova *et al.* (2020), supports the hypothesis that the motif is present in some variants. However, a crystal structure is required to confirm this.

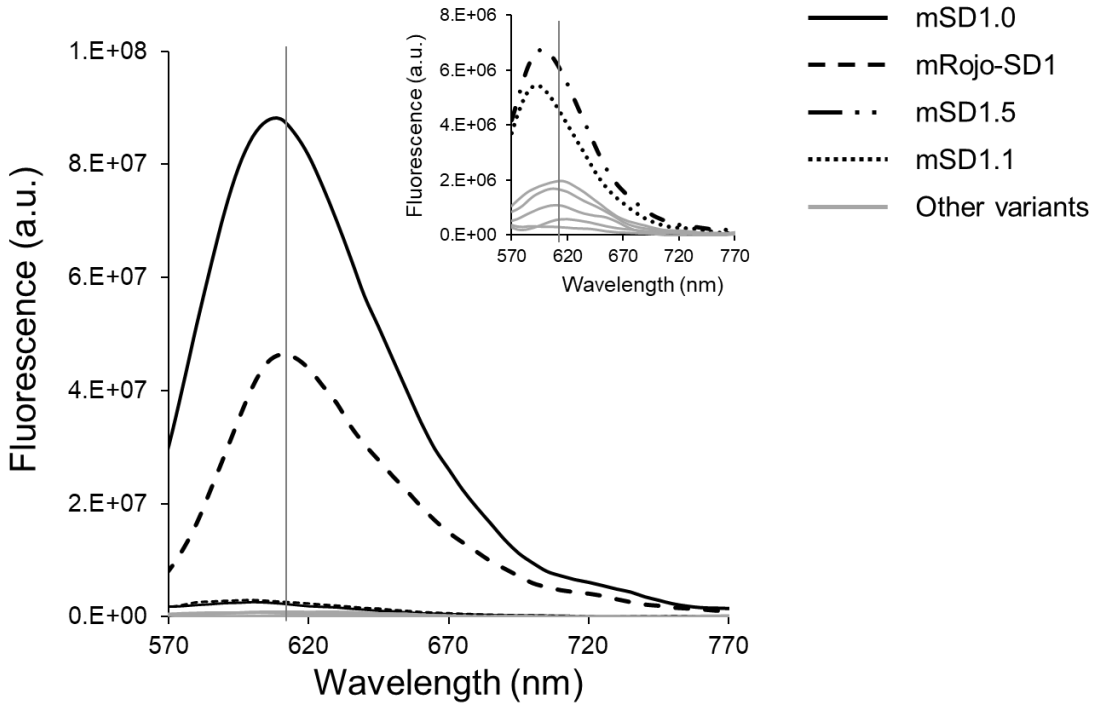
Finally, Khrenova *et al.* (2020) also predicted that π-stacking of the RFP chromophore with a tyrosine or a benzene substituent yields 9-nm or up to 21-nm bathochromic shifts, respectively[9, 16]. For most applications, red-shifted RFP variants are desired. In this work, we based our design hypothesis on the parallel π-stacking interaction found between the chromophore and a histidine in the Copepod GFP. Hereafter, future work should focus on engineering *Superdecker* motifs that instead contain a tyrosine or a phenylalanine at position 63 to create red-shifted RFPs.

2.6. Supplementary Information

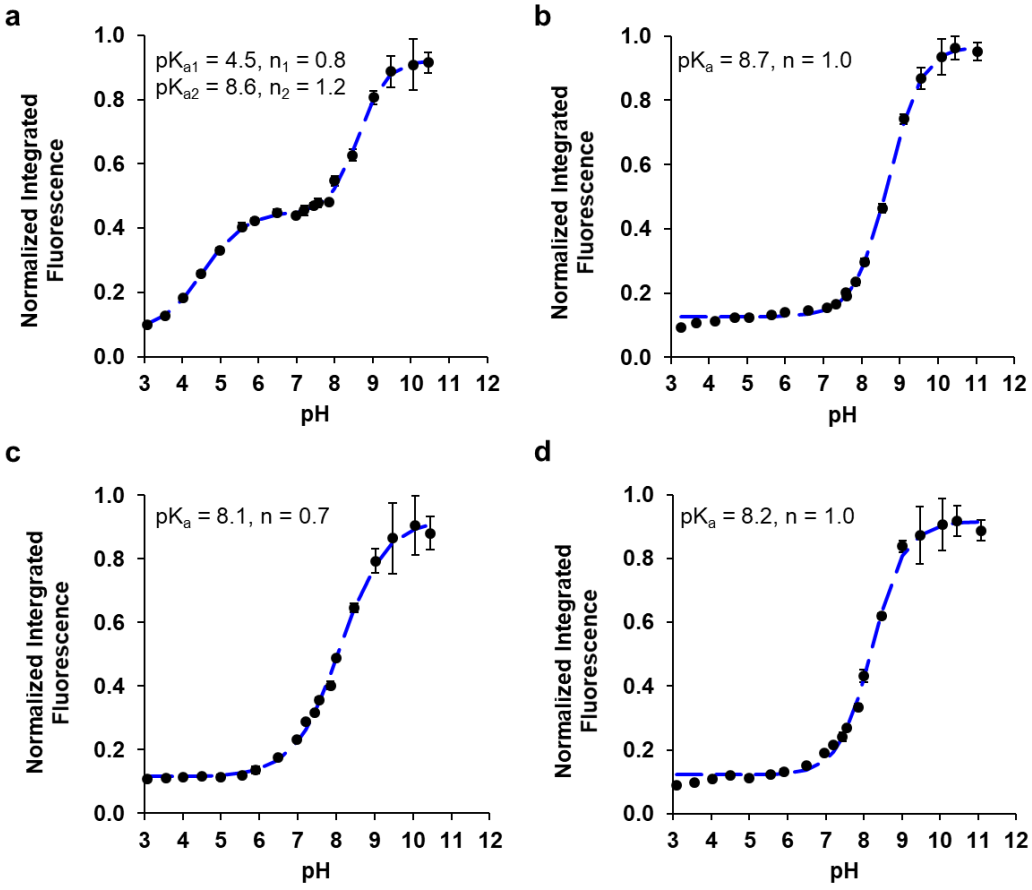
Supplementary Information includes four figures.



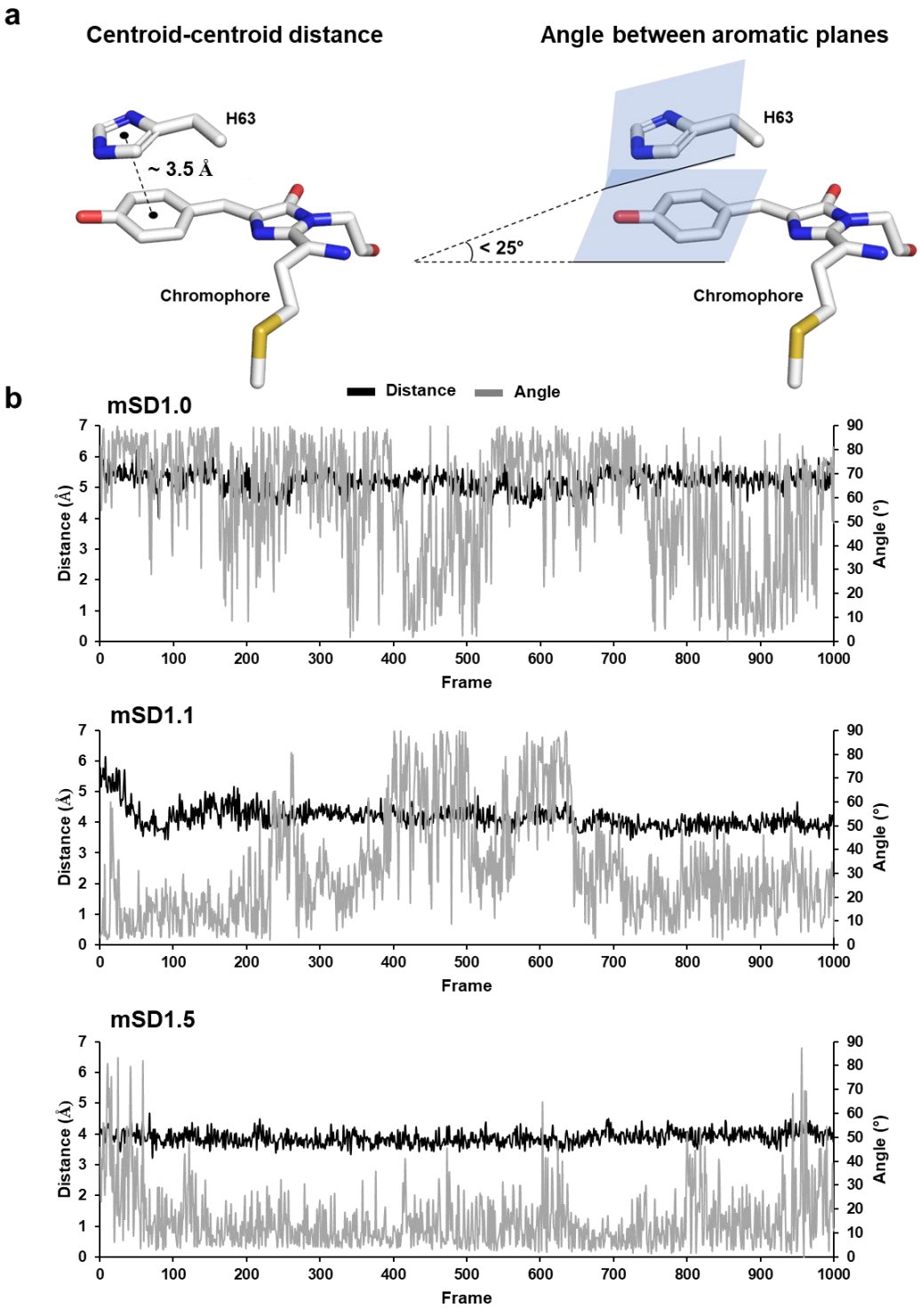
Supplementary Figure S2.1. Fast protein liquid size-exclusion chromatography (FPLC) profiles of monomeric mCherry and mRojo-SD1. mRojo-SD1 (20 mg/mL) (solid line) is a monomer, with an overlapping FPLC elution profile to that of the monomeric RFP mCherry (2 mg/mL) (dotted line). Because the mRojo-SD1-derived design mutants contain mutations focused in the protein core and not the protein surface, aside from H125R which is known to favor monomerization, we assume that these do not affect their oligomerization state and therefore are also monomeric. Proteins were purified by gel filtration into 20 mM sodium phosphate buffer pH 7.4 using an ÄKTA pure (GE Healthcare) fast protein liquid chromatography system equipped with a Superdex 75 (GE Healthcare) column.



Supplementary Figure S2.2. Experimental screening of designed *Superdecker* sequences. Emission spectra (λ_{ex} = 535 nm) of purified *Superdecker* RFPs in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4) show that two sequences display blue-shifted emission wavelengths relative to mRojo-SD1. The inset shows the emission spectra of mSD1.5 and mSD1.1 in more detail, and low brightness variants not selected for characterization. The gray vertical lines indicate the emission maximum of mRojo-SD1 (612 nm). The spectrum of mSD1.0 is shown in a black solid line, mRojo-SD1 a black dashed line, mSD1.5 a black dashed and dotted line, mSD1.1 a black dotted line, and all other designed sequences are shown as solid gray lines. Spectra were measured at room temperature using a Spectramax i3 (Molecular Devices) plate reader.

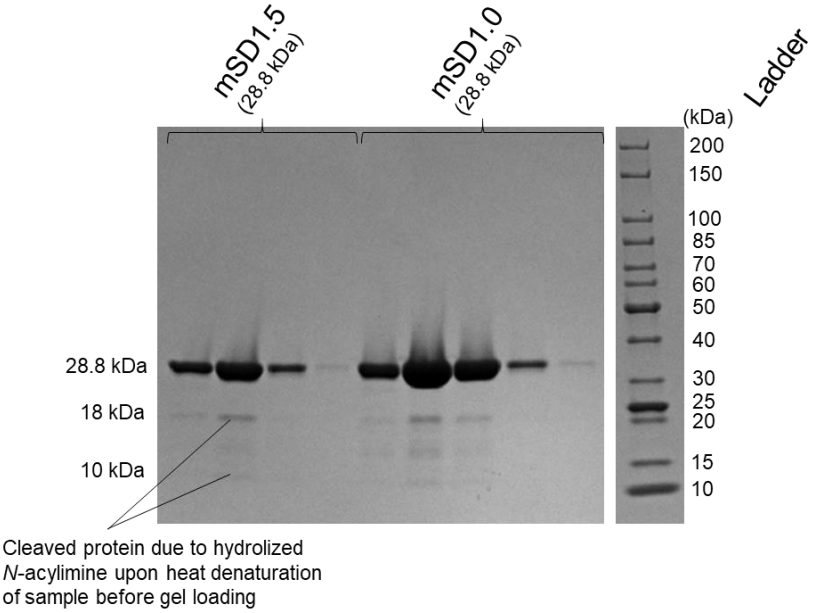


Supplementary Figure S2.3. pH profiles of *Superdecker* RFPs. pH profiles of purified (a) mRojo-SD1 (1.0 µM), (b) mSD1.0 (1.0 µM), (c) mSD1.1 (0.55 µM) and (d) mSD1.5 (1.0 µM) in a series of Britton-Robinson buffers at various pH. The pK_a is the pH value at which 50% of the maximal fluorescence is obtained, and n is the Hill's coefficient. Fluorescence intensities were measured at room temperature using a Spectramax i3 (Molecular Devices) plate reader. Data is shown as mean of triplicate ± sd.



Supplementary Figure S2.4. Molecular dynamics simulation analysis of mSD1.0, mSD1.1 and mSD1.5 *Superdecker* geometry. (a) Geometry requirements of a parallel π-stacking interaction between H63 and the phenolate moiety of the chromophore (shown as sticks). The distance between the ring centroids should be near 3.5 Å. The angle between the aromatic planes should be no higher than 25 degrees. (b) MD analysis of the

measurements described over the time course of the simulation. At the end of the simulation, 1000 snapshots separated by 0.1 ns along the production trajectory were extracted. MD analysis shows that mSD1.0 does not possess a parallel arrangement with a centroid-centroid distance around 5.5 Å and a variable angle between aromatic planes that seems to fluctuate around 90 degrees, where H63 is interestingly not in an edge-to-face orientation but is rather pointing away from the chromophore as shown in **Figure 2.1 (d)**. In contrast, mSD1.1 and mSD1.5 design models meet the requirements for a parallel H63-chromophore stacking with a distance around 4 Å and an angle below 25 degrees for the majority of the simulations. Of note, mSD1.5 seems to be the variant in which the motif may be the most stable as it retained the motif over the entire course of the simulation. The frames (x-axis) correspond to snapshots of the structure during the simulation taking at equal time intervals.



Supplementary Figure S2.5. Representative SDS-PAGE of purified Superdecker RFPs. Purified proteins in 1X SDS loading dye (0.1 M Tris-HCl pH 6.8, 0.2 M dithiothreitol, 4% w/v sodium dodecyl sulfate (SDS), 0.2% w/v bromophenol blue, 20% v/v glycerol) were heated at 95°C for 3 minutes and loaded on the gel, along with 3 µL of Unstained Protein Standard Broad Range (10-200 kDa) ladder (P7717S, NEB). The first four wells correspond to dilutions of mSD1.5 (28.8 kDa) and the next five wells correspond to dilutions of mSD1.0 (28.8 kDa). The separate well corresponds to the ladder. The additional two bands around 10 kDa and 18 kDa correspond to cleaved protein products resulting from hydrolization of the N-acylimine of the chromophore upon hear denaturation.

Supplementary Table S2.1. Crystallography data of mRojo-SD1.

	mRojo-SD1
PDB ID	TBD
Crystallization conditions	0.1 M sodium acetate pH 5.0 0.005–0.010 M $ZnCl_2$ 12% PEG-6K
Data collection [a]	
Temperature (K)	277
Resolution (Å)	1.55 – 48.04
Space group	P 31 2 1
Cell params.	
a b c (Å)	64.4499 64.4499 94.3407
α β γ (°)	90.0 90.0 120.0
Molecules per asymm. unit	1
R_{pim}	0.046 (0.719)
$CC_{1/2}$	0.996 (0.484)
I/σI	7.2 (0.9)
Completeness (%)	100 (99.4)
Multiplicity	19.1 (17.4)
# reflections (total unique)	642151 (33593)
Refinement	
R work/free	0.1518/0.1783
No. atoms	
Protein	3681
Ligand	1
Water	127
RMSD	
bond lengths (Å)	0.007
bond angles (°)	1.084
Molprobity statistics	
All-atom clashscore	2.97
Ramachandran outliers (%)	0.00
Ramachandran allowed (%)	1.43
Ramachandran favored (%)	98.57
Rotamer outliers (%)	0.00

[a] Highest resolution shell is shown in parentheses.
TBD: to be determined.

2.7. References

1. Clement, G. T.; Hynynen, K., Ultrasound phase-contrast transmission imaging of localized thermal variation and the identification of fat/tissue boundaries. *Phys Med Biol* **2005,** *50* (7), 1585-600.
2. Winnard, P. T.; Kluth, J. B.; Raman, V., Noninvasive optical tracking of red fluorescent protein-expressing cancer cells in a model of metastatic breast cancer. *Neoplasia* **2006,** *8* (10), 796-806.
3. Ash, C.; Dubec, M.; Donne, K.; Bashford, T., Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers Med Sci* **2017,** *32* (8), 1909-1918.
4. Baffour-Awuah, N. Y. A.; Zimmer, M., Hula-twisting in green fluorescent protein. *Chem Phys* **2004,** *303* (1–2), 7-11.
5. Drobizhev, M.; Molina, R. S.; Callis, P. R.; Scott, J. N.; Lambert, G. G.; Salih, A.; Shaner, N. C.; Hughes, T. E., Local Electric Field Controls Fluorescence Quantum Yield of Red and Far-Red Fluorescent Proteins. *Front Mol Biosci* **2021,** *8*, 633217.
6. Pandelieva, A. T.; Baran, M. J.; Calderini, G. F.; McCann, J. L.; Tremblay, V.; Sarvan, S.; Davey, J. A.; Couture, J. F.; Chica, R. A., Brighter Red Fluorescent Proteins by Rational Design of Triple-Decker Motif. *ACS Chem Biol* **2016,** *11* (2), 508-17.
7. Suto, K.; Masuda, H.; Takenaka, Y.; Tsuji, F. I.; Mizuno, H., Structural basis for red-shifted emission of a GFP-like protein from the marine copepod Chiridius poppei. *Genes Cells* **2009,** *14* (6), 727-37.
8. Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **2002,** *99* (12), 7877-82.
9. Khrenova, M. G.; Mulashkin, F. D.; Bulavko, E. S.; Zakharova, T. M.; Nemukhin, A. V., Dipole Moment Variation Clears Up Electronic Excitations in the π-Stacked Complexes of Fluorescent Protein Chromophores. *J Chem Inf Model* **2020,** *60* (12), 6288-6297.
10. Moore, M. M.; Oteng-Pabi, S. K.; Pandelieva, A. T.; Mayo, S. L.; Chica, R. A., Recovery of red fluorescent protein chromophore maturation deficiency through rational design. *PLoS One* **2012,** *7* (12), e52463.
11. Piatkevich, K. D.; Verkhusha, V. V., Guide to red fluorescent proteins and biosensors for flow cytometry. *Methods Cell Biol* **2011,** *102*, 431-61.
12. Cousin, S. F.; Kadeřávek, P.; Bolik-Coulon, N.; Gu, Y.; Charlier, C.; Carlier, L.; Bruschweiler-Li, L.; Marquardsen, T.; Tyburn, J. M.; Brüschweiler, R.; Ferrage, F., Time-Resolved Protein Side-Chain Motions Unraveled by High-Resolution Relaxometry and Molecular Dynamics Simulations. *J Am Chem Soc* **2018,** *140* (41), 13456-13465.
13. Riwar, L. J.; Trapp, N.; Kuhn, B.; Diederich, F., Substituent Effects in Parallel-Displaced π-π Stacking Interactions: Distance Matters. *Angew Chem Int Ed Engl* **2017,** *56* (37), 11252-11257.
14. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-47.
15. Abad, M. F.; Di Benedetto, G.; Magalhães, P. J.; Filippin, L.; Pozzan, T., Mitochondrial pH monitored by a new engineered green fluorescent protein mutant. *J Biol Chem* **2004,** *279* (12), 11521-9.
16. Khrenova, M. G.; Polyakov, I. V.; Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V., Improving the Design of the Triple-Decker Motif in Red Fluorescent Proteins. *J Phys Chem B* **2017,** *121* (47), 10602-10609.

Chapter 3: Generation of bright monomeric red fluorescent proteins via computational design of enhanced chromophore packing (Manuscript)

Sandrine Legault[1], Derek P. Fraser-Halberg[1], Ralph L. McAnelly[2], Matthew G. Eason[1], Michael C. Thompson[2] & Roberto A. Chica[1,*]

[1]Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie-Curie, Ottawa, Ontario, Canada K1N 6N5

[2]Department of Chemistry and Chemical Biology, University of California, Merced, Merced, California 95343, United States

*To whom correspondence should be addressed:

Roberto A. Chica, Email: rchica@uottawa.ca

This manuscript is to be sent for review.

3.1. Statement of Contributions

Author Contributions. R.A.C. conceived the project. R.A.C. performed computational protein design experiments. S.L. performed directed evolution experiments. S.L. and D.P.H. performed spectral characterization of RFPs. S.L., M.G.E., and D.P.H. purified proteins. R.M. and M.C.T. crystallized proteins and performed X-ray diffraction experiments. R.A.C. performed refinements. S.L. and R.A.C. wrote the manuscript. M.C.T. edited the manuscript.

Acknowledgements. R.A.C. acknowledges grants from the Canada Foundation for Innovation (26503), Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-04831), and the Human Frontier Science Program (RGP0041). S.L. is the recipient of an Ontario Graduate Scholarship and a postgraduate scholarship from the

Natural Sciences and Engineering Research Council of Canada. The authors thank Shahrokh Ghobadloo for assistance with fluorescence-activated cell sorting, Noreen Ahmed for help with protein purification, John P. Pezacki for use of the fast protein liquid chromatography system and Molecular Devices plate reader, and Stephen L. Mayo for providing the mCherry and mRojoA expression vectors.

Competing Interests

The authors declare no competing interests.

3.2. Abstract

Red fluorescent proteins (RFPs) have found widespread application in chemical and biological research due to their longer emission wavelengths. Here, we use computational protein design to increase the quantum yield and thereby brightness of a dim monomeric RFP (mRojoA, quantum yield = 0.02) by optimizing chromophore packing with aliphatic residues, which we hypothesized would reduce torsional motions causing non-radiative decay. Experimental characterization of the top 10 designed sequences yielded mSandy1 (λ_{em} = 609 nm, quantum yield = 0.26), a variant with equivalent brightness to mCherry, a widely used RFP. We next used directed evolution to further increase brightness, resulting in mSandy2 (λ_{em} = 606 nm, quantum yield = 0.35), the brightest *Discosoma* sp. derived monomeric RFP with an emission maximum above 600 nm reported to date. Crystallographic analysis of mSandy2 showed that the chromophore phenolate moiety is sandwiched between the side chains of Leu63 and Ile197, a structural motif that has not previously been observed in RFPs, and confirms that aliphatic packing leads to chromophore rigidification. Our results demonstrate that computational protein design can be used to generate bright monomeric RFPs, which could serve as templates for the evolution of novel far-red fluorescent proteins.

3.3. Introduction

Our findings in the previous chapter demonstrated that a modest increase in brightness was achieved (solely under high pH conditions) when enhancing packing of the chromophore phenolate moiety using aromatic residues. Moreover, a crystal structure is required to confirm that those observations result from the introduction of the parallel motif designed. Altogether, this suggests that using aromatic residues to enhance chromophore packing and greatly improve brightness can be challenging. Instead of aromatics, we therefore hypothesized that chromophore packing could be optimized using aliphatic residues given that their side chains can adopt various rotamers that have the ability to pack tightly within protein cores, thereby having the potential to pack tightly around the chromophore phenolate. Because the optimal combination of aliphatic residues to enhance RFP quantum yield and brightness would be difficult to predict using standard rational design methodologies (described in Chapter 1) and experimentally costly using combinatorial mutagenesis due to the large library size (~ 20^6 mutants), we decided to use a computational approach.

Here, we used computational protein design with the Triad software[10] to optimize packing of the chromophore pocket in the dim monomeric RFP mRojoA (quantum yield = 0.02), which we hypothesized would rigidify the chromophore and thereby enhance brightness. To do so, six residues surrounding the phenolate moiety of the chromophore were mutated *in silico* to various combinations of aliphatic amino acids. The best mutant identified, mSandy1, displayed an 11-fold enhancement to quantum yield relative to mRojoA. We next used directed evolution to further increase brightness, which yielded mSandy2 (quantum yield = 0.35), the brightest *Discosoma* sp. derived RFP with an

emission maximum greater than 600 nm reported to date. Crystallographic analysis confirmed that the chromophore phenolate moiety of mSandy2 was rigidified due to tight packing by aliphatic residues. Our results demonstrate the utility of computational protein design for increasing RFP brightness by enhancing chromophore packing, and generating novel templates for directed evolution of brighter variants. The approach developed here could be applied to other fluorescent proteins in order to increase quantum yield and overall brightness.

3.4. Results

Computational design of enhanced chromophore packing

Previously, we used rational design to create RFPs containing a triple-decker motif of aromatic rings in which the chromophore phenolate moiety is sandwiched between the side chains of two aromatic residues[11]. Crystallographic analysis of triple-decker RFPs showed that this motif rigidified the chromophore and resulted in modest absolute quantum yield increases of up to 0.03 at pH 7.4. Here, we postulate that chromophore packing with aliphatic (e.g., Val, Leu, Ile, and Met) instead of aromatic residues will lead to greater quantum yield increases due to the ability of bulky and/or branched aliphatic side chains to pack tightly within protein cores, which could enhance chromophore packing for more effective rigidification. However, it is difficult to predict a priori which combination of aliphatic residues will be optimal for packing the chromophore pocket, emphasizing the need for a computational approach.

Algorithms for computational protein design (CPD) optimize amino-acid side-chain rotamers on a backbone template to stabilize a protein fold by improving non-bonding interactions[12, 13]. These methods can therefore be used to predict combinations of aliphatic residues to enhance packing of the chromophore phenolate moiety without destabilizing the protein fold. CPD has successfully been used to repack protein cores for enhanced stability, often with high accuracy[14, 15]. CPD has also been used to design RFPs displaying red-shifted emission wavelengths[16] and create monomeric variants from oligomeric ones[17, 18]. However, it has not, to the best of our knowledge, been applied to the repacking of the RFP chromophore pocket for enhanced brightness.

As a starting template for design, we selected mRojoA, a dim and red-shifted variant of the widely-used monomeric RFP mCherry that we engineered for red-shifted fluorescence[16]. mRojoA contains the V16T, R125H, Q163L, V195A, I197Y, and A217C mutations (numbering according to the mCherry structure, PDB ID: 2H5Q[19]) that together cause a 23-nm emission wavelength bathochromic shift that is accompanied by an 10-fold decrease in quantum yield (**Table 3.1**). We selected mRojoA as our design template because it is dim, its crystal structure is available (PDB ID: 3NEZ[16]), and it was the template used to create the triple-decker RFPs[11], allowing us to compare the structural and functional effects of aliphatic vs. aromatic packing of the chromophore. Furthermore, the mRojoA crystal structure shows suboptimal packing around the chromophore due to the presence of a Pro residue at position 63 (**Figure 3.1**), whose cyclic side chain cannot form tight packing interactions against the planar p-hydroxybenzylidene moiety. Mutation of this Pro residue, which is conserved among all *Discosoma* sp. derived RFPs[20], to non-cyclic aliphatic amino acids should enable us to test our design hypothesis.

Using an ensemble of four mRojoA templates (see **3.7. Materials and Methods**), we designed six first-shell residues surrounding the chromophore (**Supplementary Figure S3.1a**) and allowed only aliphatic residues (Ile, Leu, Met, Val) at these positions with the exception of position 143, where we also allowed Ser because this amino acid has been shown to contribute to enhanced brightness in triple-decker RFPs[11]. We also reverted the V16T mutation of mRojoA because polar residues at this position have been shown to hinder red chromophore maturation[21], and the R125H surface mutation to enhance solubility in the monomeric state[4]. The top 10 designed sequences (**Supplementary Figure S3.1b**) contained between 4 and 7 mutations, including P63L directly on top of the chromophore, and W143S, which is likely required to accommodate the bulky Leu63 side chain. At the other four design positions, various combinations of aliphatic residues were observed in the top 10 ranked sequences, which were selected for experimental characterization.

Table 3.1. Spectral properties of red fluorescent proteins.

Protein	λ_{ex} (nm)	λ_{em} (nm)	Quantum Yield	Extinction Coefficient ($mM^{-1} cm^{-1}$)	Brightness ($mM^{-1}cm^{-1}$)	Mutations relative to mRojoA												
						T16	P63	H125	W143	I161	L163	L165	D174	N194	A195	Y197	L199	C217
mRojoA	596 ± 2	632 ± 2	0.023 ± 0.001	68 ± 9	2	-	-	-	-	-	-	-	-	-	-	-	-	-
mRojoA-Y197I	588 ± 1	618 ± 1	0.076 ± 0.003	75 ± 2	-	-	-	-	-	-	-	-	-	-	-	I	-	-
mSandy0.1	585	608	0.13	-	-	V	L	R	S	V	-	-	-	-	-	V	-	-
mSandy0.7	584	612	0.16	94	15	V	L	R	S	-	-	-	-	-	-	V	-	-
mSandy0.9	585	608	0.13	-	-	V	L	R	S	V	V	-	-	-	-	V	-	-
mSandy1	584 ± 1	609 ± 2	0.26 ± 0.02	81 ± 5	21	V	L	R	S	V	-	-	-	-	-	I	-	-
mSandy2	581 ± 1	606 ± 1	0.35 ± 0.03	79 ± 9	28	V	L	-	S	V	V	F	V	Y	-	I	M	-
mSandy1-L163V/L199M	580	606	0.28	-	-	V	L	R	S	V	V	-	-	-	-	I	M	-
mCherry	586 ± 2	609 ± 1	0.23 ± 0.01	87 ± 3	20	V	-	R	-	-	Q	-	-	-	V	I	-	A

λex and λem correspond to the excitation and emission maxima, respectively.
n ≥ 3 independent experiments for mRojoA, mRojoA-Y197I, mSandy1, mSandy2, and mCherry (mean ± s.d). n = 1 independent experiment for all other variants. Each independent experiment consisted of triplicate measurements performed on a single protein batch.
All values are at pH 7.4 in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate).
Brightness is the product of the quantum yield and extinction coefficient.

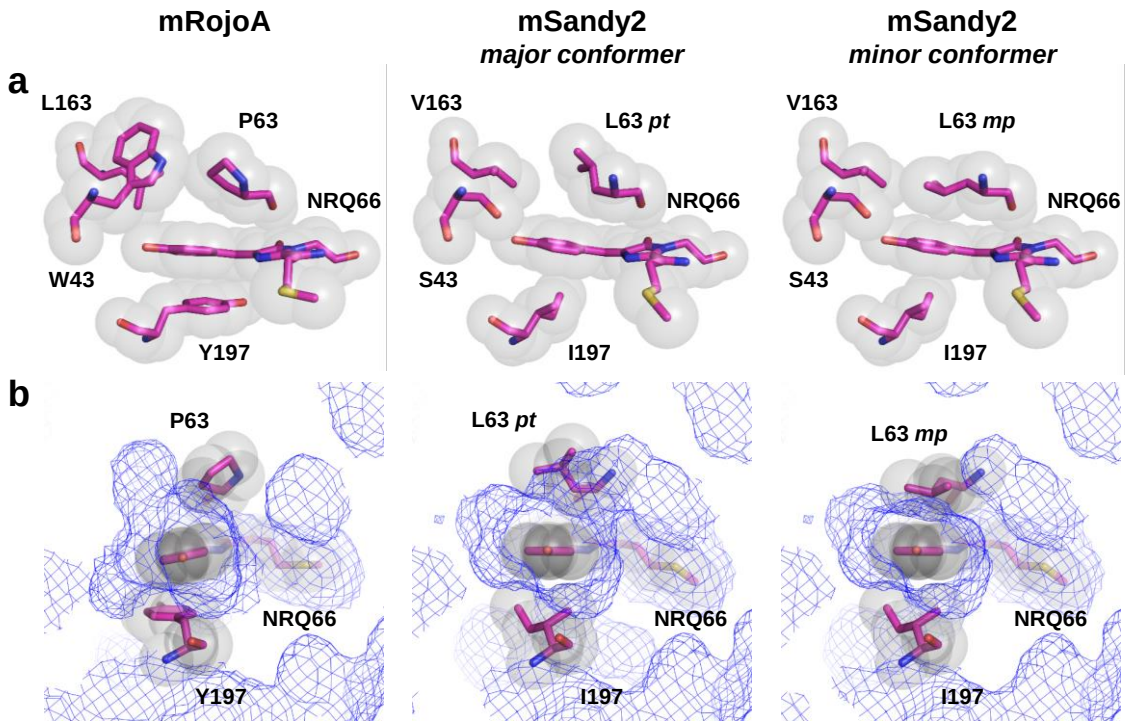


Figure 3.1. RFP chromophore pocket. Key residues forming part the chromophore (NRQ66) pocket in mRojoA (chain A, PDB ID: 3NEZ) and mSandy2 (chain H, PDB ID: unpublished) are shown as sticks and transparent spheres. (a) The chromophore pocket in mRojoA contains a cavity above the chromophore phenolate moiety due to the cyclic structure of P63. This cavity is filled in mSandy2 by aliphatic residue L63 when it adopts the *mp* but not *pt* rotamer. Occupancy of the *mp* rotamer varies between 30–53% in the 8 molecules of the crystal asymmetric unit. (b) Blue mesh shows cavity surrounding the chromophore phenolate moiety in mRojoA and the mSandy2 major conformer. This cavity is filled by the L63 *mp* rotamer. All images generated using PyMOL 2.5.0.

Spectral characterization of designs

Four of the 10 designed sequences showed strong fluorescence with emission maxima around 610 nm (**Supplementary Figure S3.2**), which represent hypsochromic shifts of approximately 20 nm relative to mRojoA. These results can be partially ascribed to reversion of the V16T and I197Y mutations that are known to cause emission wavelength bathochromic shifts of 4 and 7 nm, respectively, when introduced into mCherry[16]. These variants, which we call mSandys (**Table 3.1**, **Figure 3.2**,

Supplementary Figure S3.3), all display quantum yield increases of 5.6–11-fold relative to mRojoA, representing absolute quantum yield increases of 0.11–0.23. The brightest variant, mSandy1, is as bright as mCherry with a quantum yield of 0.26 and a brightness of 21 mM^{-1} cm^{-1}. mSandy1 also displays efficient red chromophore maturation, with no apparent green chromophore peak (λ = 510 nm, **Supplementary Figure S3.4**) in its absorption spectrum (**Figure 3.2**), similar to what is observed for the efficiently-maturing mCherry. Importantly, these comparable properties of mSandy1 and mCherry are not simply due to a high sequence similarity between these two RFPs. As shown on **Table 3.1**, the sequence of mSandy1 is more similar to that of its parent mRojoA (6 mutations) than it is to that of mCherry (7 mutations). This result highlights how different sequence changes can produce similar functional effects in RFPs.

Of the bright mSandy variants, mSandy1 is the only one that contains the Y197I mutation, which reverts the Tyr found at this position in mRojoA to the corresponding Ile found in mCherry. Since the single point mutation I197Y causes a large quantum yield decrease of 0.19 when introduced into mCherry[16], we verified whether the Y197I reversion found in mSandy1 was primarily responsible for the large quantum yield increase observed in this variant. As shown on **Table 3.1**, the single mutant mRojoA-Y197I displays an absolute quantum yield increase relative to mRojoA of approximately 0.05, less than that observed for mSandy1 (0.23). This result confirms that presence of the Y197I mutation alone is not sufficient to account for the high brightness of mSandy1, and that the other five mutations contribute synergistically to its high brightness.

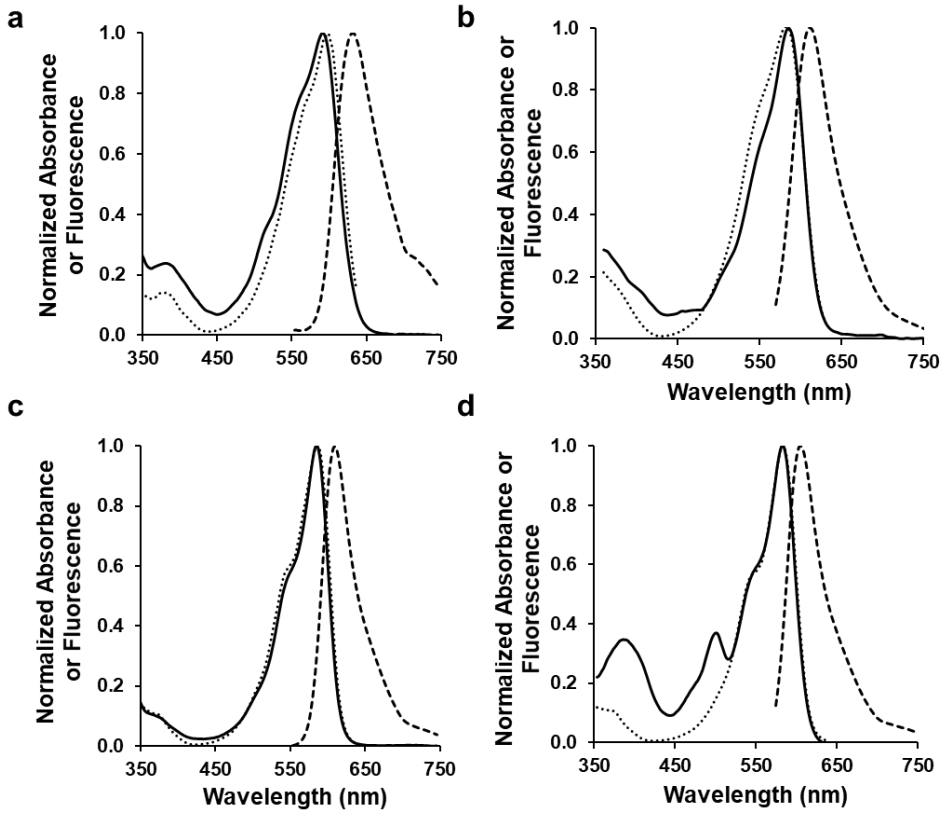


Figure 3.2. Fluorescence and absorption spectra. (a) mRojoA, (b) mSandy1, (c) mCherry, and (d) mSandy2. Absorption, excitation, and emission spectra are shown as solid, dotted, and dashed lines, respectively. Spectra of purified proteins were measured in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4) at room temperate using an Infinite M1000 (Tecan) or Spectramax i3 (Molecular Devices) plate reader.

Evolution of mSandy2

Given that mSandy1 is as bright as mCherry but contains an amino acid at position 63 (Leu63) that has never been observed in other fluorescent proteins (**Supplementary Data File**), we postulated that it could serve as a good starting template for the directed evolution of brighter monomeric RFPs. We therefore subjected mSandy1 to random mutagenesis followed by high-throughput screening for enhanced brightness. After three rounds of directed evolution (see **3.7. Materials and Methods**), we isolated mSandy2

(**Table 3.1**, **Figure 3.2**), a variant whose emission wavelength is not significantly altered but that displays a quantum yield that is 35% higher than that of mSandy1 (absolute quantum yield increase of 0.09). mSandy2 is also 40% brighter than mCherry, an RFP that emits at a similar wavelength. However, chromophore maturation in mSandy2 is less efficient than in either mSandy1 or mCherry, as evidenced by the peaks at approximately 390 and 510 nm in its absorption spectrum (**Figure 3.2d**) that correspond to neutral and anionic green chromophores (**Supplementary Figure S3.4**), respectively[21, 22]. Presence of these absorption peaks indicates a substantial amount of green chromophore-containing molecules in the RFP population in solution.

mSandy2 contains six mutations from mSandy1, including two that occurred at designed positions (L163V and L199M). To evaluate whether these two mutations were sufficient to account for the large quantum yield increase in this variant, we generated the mSandy1-L163V/L199M double mutant. These two mutations together improve the quantum yield of mSandy1 by 0.02 (**Table 3.1**), less than half of the total quantum yield increase in mSandy2 relative to mSandy1. Furthermore, they are detrimental to chromophore maturation, as evidenced by the low overall absorption and strong peaks at approximately 390 and 510 nm (**Supplementary Figure S3.3**). The other four mutations therefore contribute to increase the quantum yield and compensate for the maturation deficiency caused by the L163V/L199M combination. Interestingly, three of the remaining four mutations occur on the RFP β-barrel surface, including a reversion of the H125R mutation that we introduced to favor monomer solubility. Nevertheless, these surface mutations do not cause oligomerization of mSandy2, which remains a monomer in solution (**Supplementary Figure S3.5**).

Crystal structure of mSandy2

We next solved the crystal structure of mSandy2 to evaluate the structural factors contributing to its high quantum yield. Crystals of mSandy2 were obtained at pH 8.5, and these diffracted at a resolution of 2.05 Å (**Table 3.2**). The unit cell corresponded to space group $P2_12_12_1$ with eight molecules in the asymmetric unit. Interestingly, only two of these molecules (chains E and H) contain red chromophores, with all others containing green chromophores in which the *N*-acylimine substituent of the chromophore is not formed. This result is in agreement with the observed mixture of green and red chromophore-containing mSandy2 molecules in solution (**Figure 3.2**). Nevertheless, in all eight molecules within the asymmetric unit, the phenolate moiety of the chromophore is sandwiched between the isobutyl and *sec*-butyl side chains of Leu63 and Ile197, a structural motif that has not been previously observed in fluorescent proteins. In these structures, the Leu63 side chain adopts two distinct rotamers (**Figure 3.1**). The *mp* rotamer (χ_1 = gauche(–), χ_2 = gauche(+)), which is found at this position in our computational model of mSandy1 (**Supplementary Figure S3.6**), fills the cavity above the phenolate moiety of the chromophore by packing tightly against it. In contrast, the alternate *pt* rotamer (χ_1 = gauche(+), χ_2 = trans), does not. Interestingly, the *mp* rotamer of Leu63 is the minor conformer in all but one of the mSandy2 molecules in the crystal asymmetric unit, with occupancies between 30% and 53%. Nevertheless, the Leu63 *mp* rotamer likely contributes to reduce non-radiative bond rotation within the chromophore p-hydroxybenzylidene moiety, leading to enhanced quantum yield.

Table 3.2. Crystallographic data and refinement statistics

	mSandy2
PDB ID	TBD
Crystallization conditions	100 mM Tris buffer (pH 8.5) 100 mM NaCl 20% PEG-3350 34 mg/mL protein
Data collection [a]	
Resolution (Å)	80.31–2.05
Space group	$P2_12_12_1$
Cell params.	
a b c (Å)	59.68 147.21 240.94
α β γ (°)	90 90 90
Molecules per asymm. unit	8
R_{pim}	0.052 (0.890)
$CC_{1/2}$	0.999 (0.396)
I/σI	11.9 (1.0)
Completeness (%)	99.9 (97.3)
Multiplicity	13.0 (10.9)
# unique reflections	133939
Refinement	
R work/free	0.1823/0.2255
No. atoms	
Protein	14089
Water	641
RMSD	
bond lengths (Å)	0.007
bond angles (°)	0.958
Molprobity statistics	
All-atom clashscore	3.09
Ramachandran outliers (%)	0.00
Ramachandran allowed (%)	1.78
Ramachandran favored (%)	98.22
Rotamer outliers (%)	0.47

[a] Highest resolution shell is shown in parentheses.
TBD: to be determined.

To verify whether aliphatic packing mutations increase chromophore rigidity, we analyzed B-factors of individual chromophore atoms, which correspond to their average displacement in the crystal. Since both conformational flexibility and crystalline disorder can contribute to atomic B-factors, we calculated their Z-scores to account for variations

between different crystals and molecules, and compared those of mSandy2 with those of mRojoA and mCherry. B-factor Z-scores for the chromophore phenolate moiety are lower in mSandy2 than in its dim parent mRojoA (**Figure 3.3**, **Supplementary Figure S3.7**), confirming that the mSandy2 chromophore is more rigid. The mSandy2 chromophore phenolate is also more rigid than that of mCherry, which is itself more rigid than that of mRojoA. Taken together, these results demonstrate that as rigidity of the chromophore phenolate increases, so does RFP quantum yield. Overall, our data confirm that the computationally-designed aliphatic packing mutations, in combination with those found by directed evolution, cause phenolate rigidification leading to quantum yield increases.

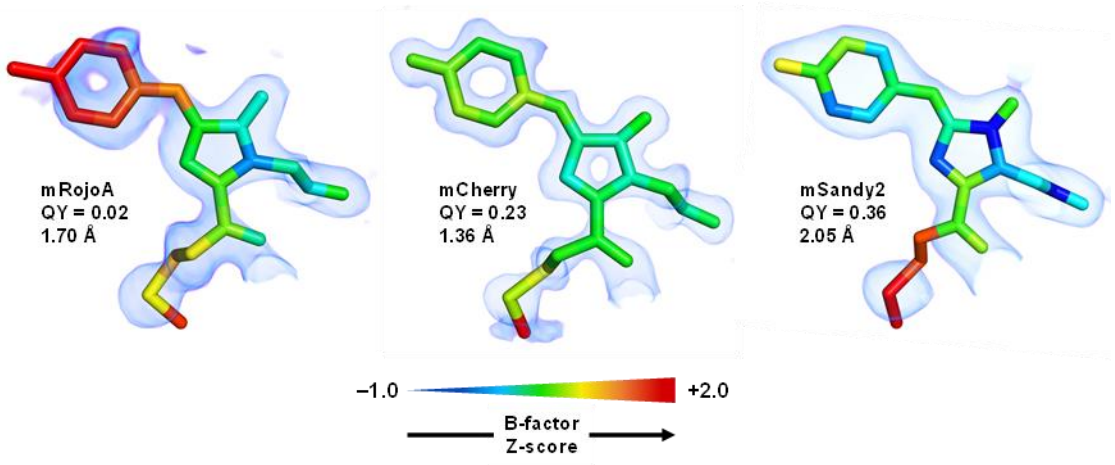


Figure 3.3. RFP quantum yield increases alongside chromophore rigidity. The red chromophores of mRojoA (chain A, PDB ID: 3NEZ), mCherry (PDB ID: 2H5Q), and mSandy2 (chain E, PDB ID: unpublished) are shown as sticks with individual atoms colored according to their B-factor Z-score. The 2Fo-Fc map is shown in volume representation at a contour level of 1.0σ in light blue. The crystal structure resolution and quantum yield (QY) are indicated for each protein.

3.5. Discussion

In this work, we used computational protein design to optimize packing of the RFP chromophore with aliphatic residues to reduce its conformational flexibility and thereby enhance its quantum yield and overall brightness. In the brightest computationally designed variant, mSandy1, aliphatic packing of the chromophore causes an absolute quantum yield increase of 0.23 relative to the mRojoA parent without significantly affecting the molar extinction coefficient. This results in a brightness increase of approximately one order of magnitude. The quantum yield increase obtained by aliphatic packing described here is larger than those obtained by aromatic packing in the triple-decker RFPs. For example, the brightest triple-decker RFP, mRojo-VYGV, which contains a chromophore phenolate moiety sandwiched between the side chains of Tyr63 and Tyr197, displays an absolute quantum yield increase of 0.03 and a 2-fold increase in brightness relative to mRojoA at pH 7.4[11]. The approximately 5-fold higher quantum yield of mSandy1 (0.26) compared to mRojo-VYGV (0.05) suggests that aliphatic packing is better suited for increasing RFP brightness than aromatic packing, in agreement with our design hypothesis. Importantly, our computational design approach required the testing of only 10 sequences, which is many orders of magnitude lower than the number of sequences that are typically required to achieve comparable increases in quantum yield by directed evolution[4, 5]. These results highlight the benefit of CPD, which is to facilitate the identification of mutation combinations that would have been difficult to predict rationally or to obtain through random mutagenesis[16].

Directed evolution on the computationally designed mSandy1 yielded mSandy2, which is, to the best of our knowledge, the brightest *Discosoma* sp. derived monomeric

RFP with an emission maximum above 600 nm reported to date. The quantum yield of mSandy2 (0.35) is comparable to those of mRuby (0.35) and mKate2 (0.40), two monomeric RFPs that have been used in many imaging applications[23, 24]. mRuby emits at a similar wavelength as mSandy2 (λ_{em} = 605 nm) but is 40% brighter (39.2 mM^{-1} cm^{-1}) due to a higher reported extinction coefficient. On the other hand, mKate2 is red-shifted by 27 nm but is 12% less bright than mSandy2 (25 mM^{-1} cm^{-1}). Interestingly, mRuby and mKate2 were engineered from natural RFPs isolated from the *Entacmaea quadricolor* sea anemone (eqFP611 and eqFP578, respectively), and both contain a Thr residue at position 63. In the crystal structures of these proteins, the Thr63 side-chain methyl group packs against the chromophore phenolate moiety, likely contributing to chromophore rigidification. While several monomeric far-red fluorescent proteins (λ_{em} >650 nm) with quantum yields ≥0.10 have been evolved from eqFP578 and eqFP611 (e.g., mKelly1, mMaroon1, and mGarnet), to our knowledge, none have been successfully engineered from either DsRed[25] or HcRed[26], and only relatively dim variants (quantum yield ≤0.10) with emission maxima <650 nm have been produced (e.g., mRouge[16], mPlum[27], mGinger1[18]. Presence of Thr at position 63 instead of a Pro could partially explain why RFPs from *Entacmaea quadricolor* have been better templates for evolving far-red fluorescent proteins than those from *Discosoma* sp. (DsRed[25]) or *Heteractis crispa* (HcRed[26]), which instead contain a Pro residue at the corresponding position. Since RFP quantum yield tends to decrease as emission wavelength increases, we propose that the engineering of novel far-red fluorescent proteins could be initiated from a template where the chromophore p-phenolate moiety is already tightly packed. Our mSandy variants could be useful for this purpose.

Although mSandy2 is bright, further improvements to its quantum yield should be possible. This could be achieved by using CPD to design second-shell mutations to stabilize the Leu63 side chain in the *mp* conformation that allows it to pack tightly against the chromophore phenolate. In addition to helping to rigidify the chromophore, mutations introduced in the second shell should be less likely to affect the emission wavelength, as they are not in direct contact with the chromophore. The aliphatic packing approach described here could also be applied to increase the brightness in other Pro63-containing RFPs such as mGinger1 (λ_{em} = 637 nm, quantum yield = 0.02)[18] or mPlum (λ_{em} = 649 nm, quantum yield = 0.10)[21], and in near-infrared fluorescent proteins derived from bacterial phytochromes[28] by helping to improve packing of the biliverdin chromophore for increased rigidity. While many RFPs are now available, novel variants displaying high brightness in the far-red or near-infrared optical window, as well as superior photostability and fast maturation, are still desired to enable various types of applications. In this context, the approach combining CPD with directed evolution described here has potential to yield novel templates for further engineering.

3.6. Supplementary Information

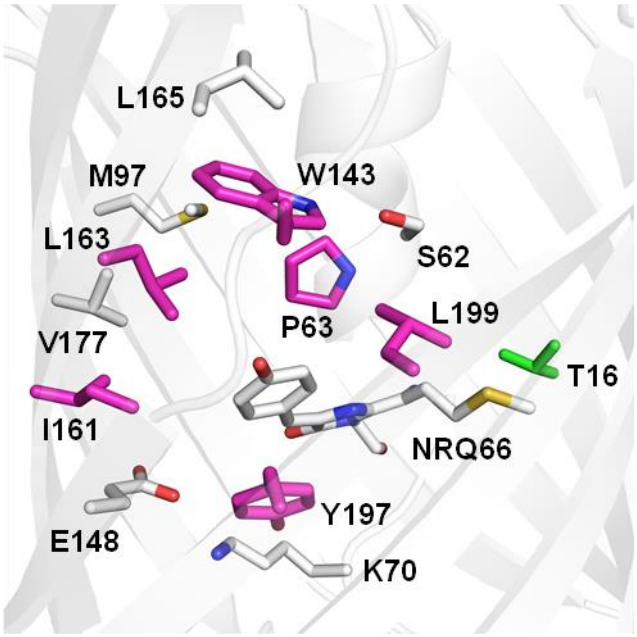
Supplementary Information includes one data file, seven figures, and two tables.

Supplementary Data File.

https://www.ncbi.nlm.nih.gov/projects/msaviewer/?anchor=0&coloring=diff&key=VebPP0nklj06ytg6GdvuxL5ySugVmRucF5o_jCuIOaaoq-JDtHmzOmCfVTdOgM_9nuXD8d3VhtDBytXH08Hf2u3z0P38wdY,ZtX8DHrXpQ4J-esJKujd941Bak41Pzs6NzwfKgsuGQCIDcLllN-yzfjOzXVJNkpLG1NGR1hjA2ZEfFBxVndabGhFVUt5d1M&track_config=protein_default&from=30&to=90&columns=d:120,b:55,x:17,aln,e:55,o:150

The file shows the amino acid sequence of mSandy1 aligned to similar sequences using the Multiple Sequence Alignment Viewer 1.20.0 (NCBI). The alignment was generated using Blastp (protein-protein BLAST), in which mSandy1 was entered as the query sequence and was "blasted" using the non-redundant protein sequences (nr) database, against a maximum of 1000 target sequences. The aligned sequences were then filtered to exclusively include those with accession lengths of 200-250 amino acids, which correspond to the expected size of fluorescent proteins from the GFP superfamily. To focus on residue L63 of mSandy1, the file shows sequences from residue 30 to residue 90. Differences are highlighted in red. The multiple sequence alignment shows that a Leu at position 63 is not found in other FPs.

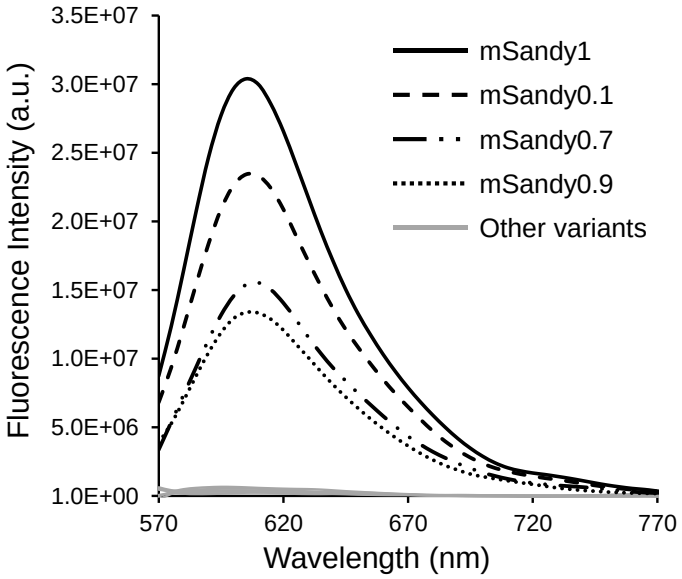
a



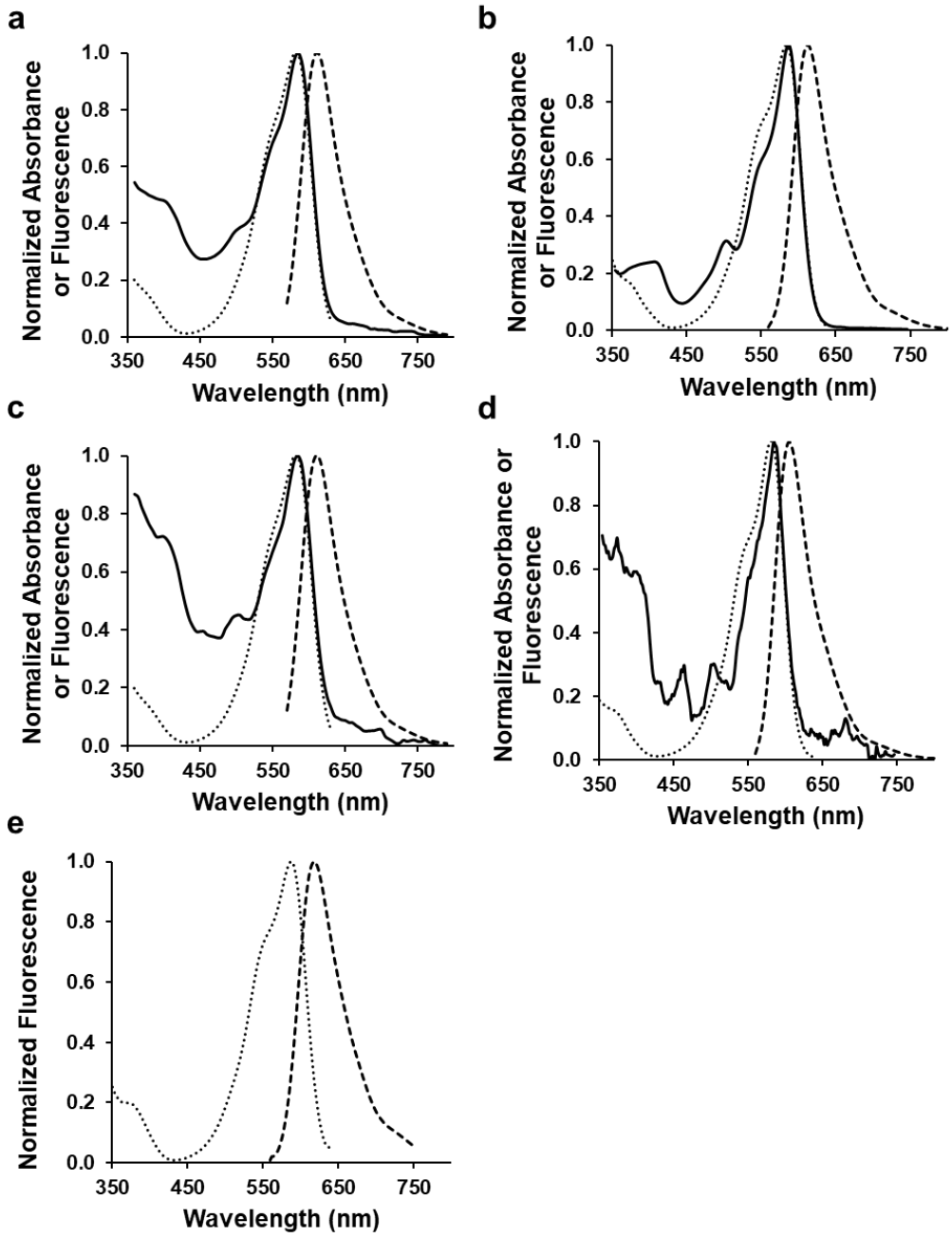
b

		Positions													
Seq	Energy	16	62	63	66	70	97	143	148	161	163	165	177	197	199
WT		T	S	P	NRQ	K	M	W	E	I	L	L	V	Y	L
0	-2342.436566	V	-	L	*	-	-	S	-	V	-	-	-	V	I
1	-2341.211358	V	-	L	*	-	-	S	-	V	-	-	-	V	-
2	-2340.200948	V	-	L	*	-	-	S	-	V	-	-	-	I	I
3	-2338.936610	V	-	L	*	-	-	S	-	V	-	-	-	I	-
4	-2338.836003	V	-	L	*	-	-	S	-	V	V	-	-	V	I
5	-2338.634281	V	-	L	*	-	-	S	-	V	-	-	-	V	V
6	-2338.172567	V	-	L	*	-	-	S	-	-	-	-	-	V	I
7	-2336.947382	V	-	L	*	-	-	S	-	-	-	-	-	V	-
8	-2336.675792	V	-	L	*	-	-	S	-	V	V	-	-	I	I
9	-2336.586179	V	-	L	*	-	-	S	-	V	V	-	-	V	-

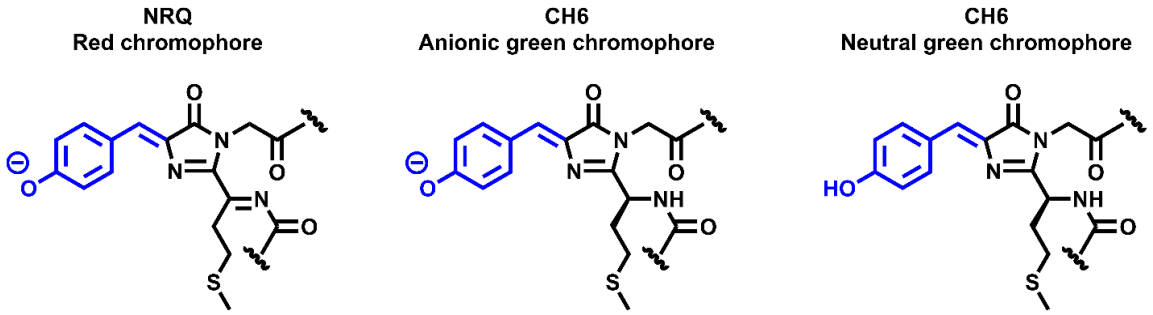
Supplementary Figure S3.1. Computational design of mSandy1. (a) Structure of mRojoA (PDB ID: 3NEZ) with designed residues shown as sticks and identified by their one-letter code. NRQ indicates the chromophore. Residues whose identify and conformation were allowed to change are colored magenta, while those whose conformation but not identity was allowed to change are colored white. The T16V mutation introduced to facilitate chromophore maturation is colored green. (b) Multistate design results. Asterisks, letters and dashes indicate the chromophore, mutations, or wild-type residues, respectively. The top 10 sequences, ranked according to their Boltzmann-weighted average energy across the four-template ensemble, were experimentally characterized. Sequence 3 corresponds to mSandy1.



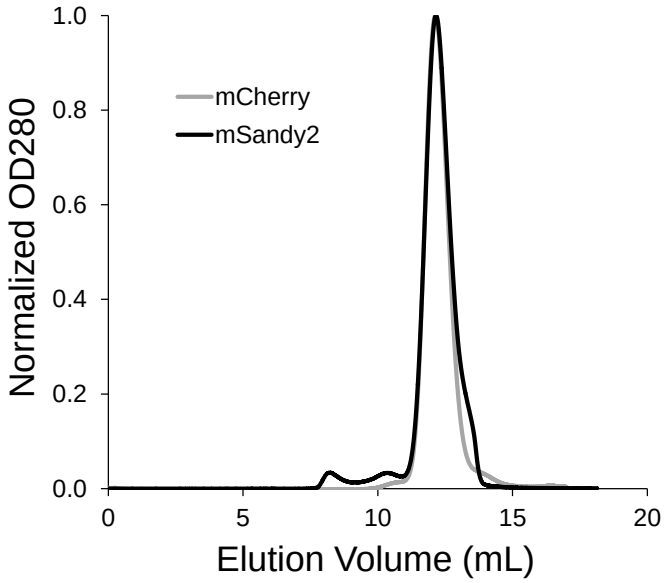
Supplementary Figure S3.2. Screening of designed sequences for bright fluorescence. Emission spectra (λ_{ex} = 535 nm) of purified samples of the top 10 designed mSandys in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4) revealed that only four of these variants displayed high fluorescence. Spectra were measured at room temperature using a Spectramax i3 (Molecular Devices) plate reader.



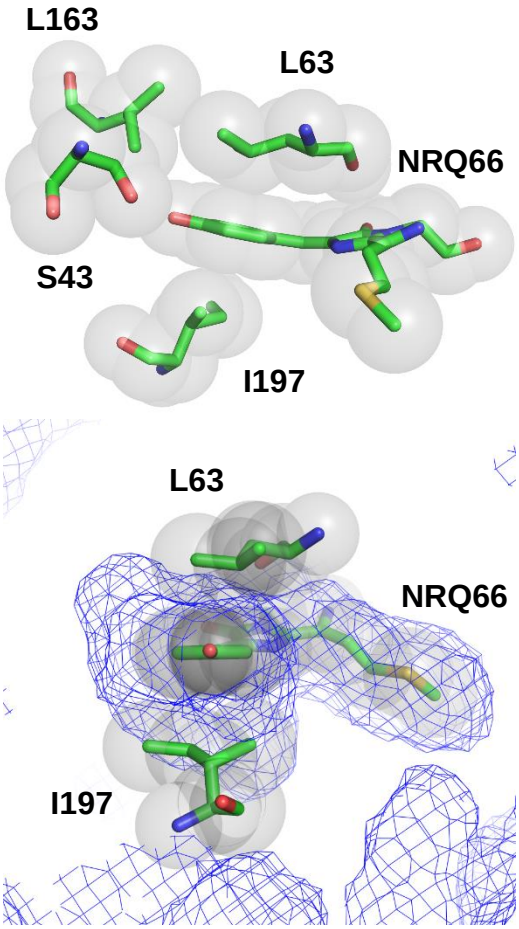
Supplementary Figure S3.3. Fluorescence and absorption spectra of mSandy variants. (a) mSandy0.1, (b) mSandy0.7, (c) mSandy0.9, (d) mSandy1-L163V/L199M, and (e) mRojoA-Y197I. Absorption, excitation, and emission spectra are shown as solid, dotted, and dashed lines, respectively. Spectra of purified proteins were measured in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4) at room temperate using a Spectramax i3 (Molecular Devices) plate reader.



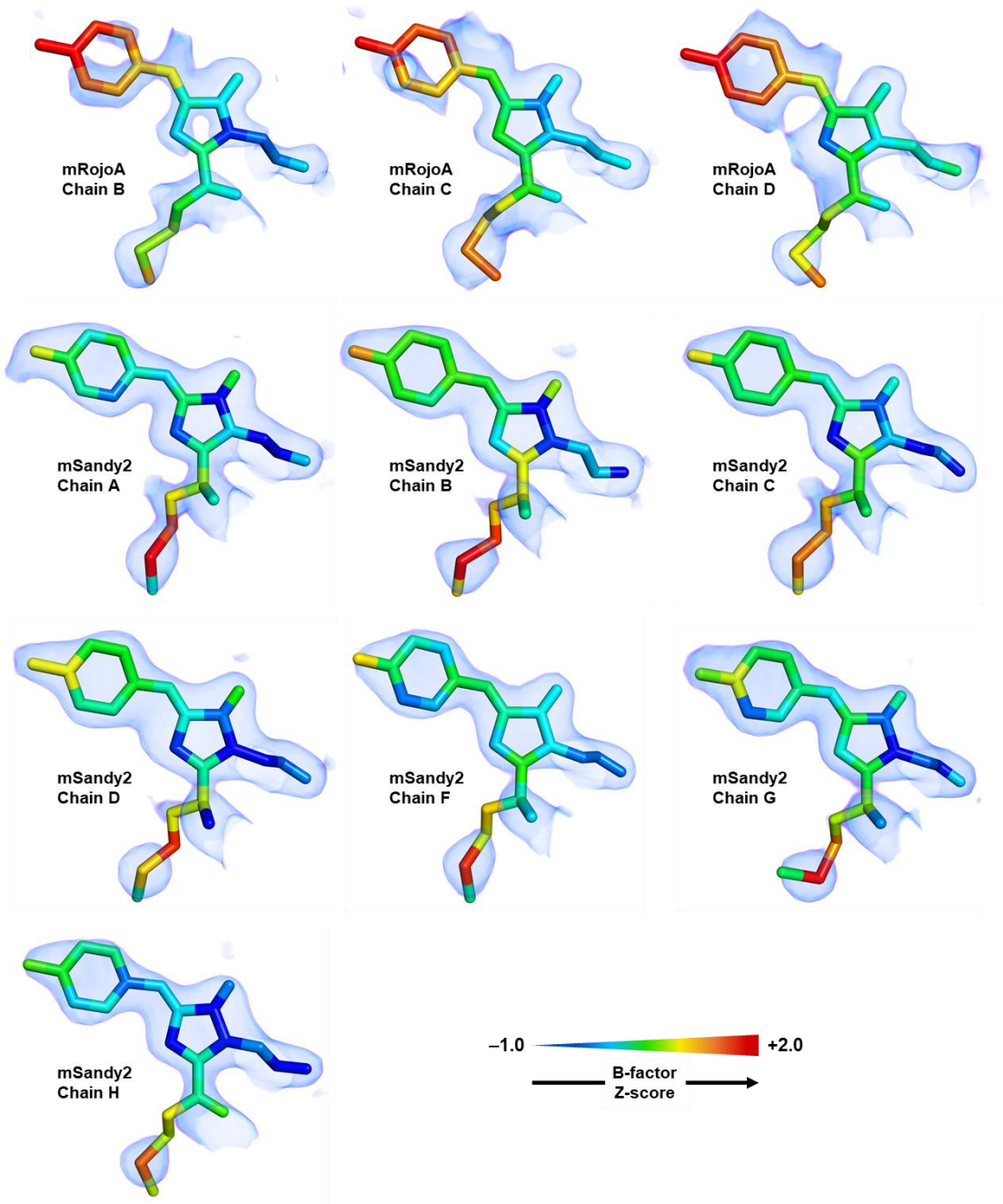
Supplementary Figure S3.4. Chromophore structure. The red chromophore differs from the green one by the presence on an *N*-acylimine group that extends the size of the conjugated system. Red, anionic green, and neutral green chromophores absorb at approximately 590 nm, 510 nm, and 390 nm, respectively. The p-hydroxybenzylidene (or phenolate) moiety of each chromophore is colored blue. Ask Roberto of need to remove this figure.



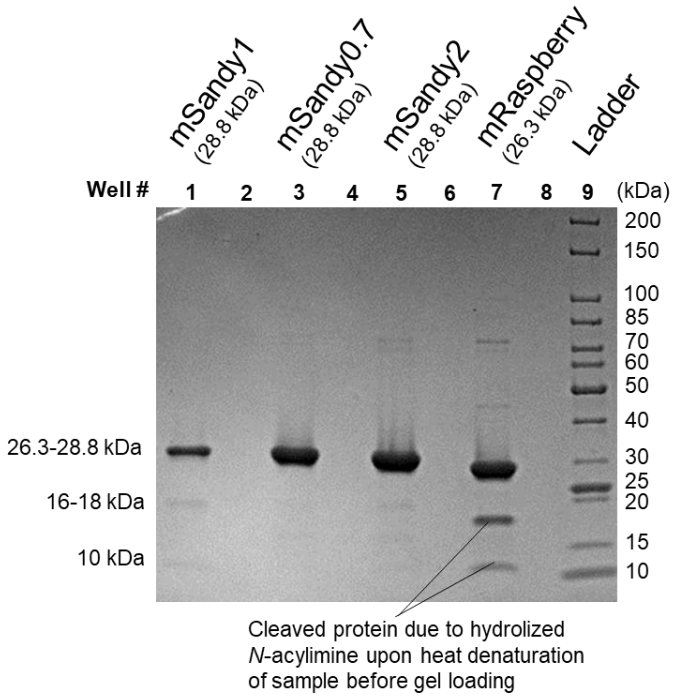
Supplementary Figure S3.5. mSandy2 is a monomeric red fluorescent protein. Size-exclusion chromatography elution profiles monitored by absorbance at 280 nm during the final purification step show a single predominant peak for mSandy2 (25 mg/mL), which elutes at the same volume as mCherry (2 mg/mL), a known monomeric red fluorescent protein. Proteins were purified by gel filtration into 20 mM sodium phosphate buffer pH 7.4 using an ÄKTA pure (GE Healthcare) fast protein liquid chromatography system equipped with a Superdex 75 (GE Healthcare) column.



Supplementary Figure S3.6. mSandy1 computational model. Key residues forming part the chromophore (NRQ66) pocket are shown as sticks and transparent spheres (a) The designed Leu63 residue adopts the mp rotamer and tightly packs against the chromophore phenolate moiety. (b) Blue mesh shows chromophore pocket cavity. All images generated using PyMOL 2.5.0.



Supplementary Figure S3.7. Chromophore rigidity. The chromophores of mRojoA (PDB ID: 3nez) and mSandy2 (PDB ID: unpublished) are shown as sticks with atoms colored according to their B-factor Z-score. The 2Fo-Fc map is shown in volume representation at a contour level of 1.0σ. mRojoA Chain A and mSandy2 Chain E are shown in **Figure 3.3**.



Supplementary Figure S3.8. Representative SDS-PAGE of purified mSandys and mRaspberry control. Purified proteins in 1X SDS loading dye (0.1 M Tris-HCl pH 6.8, 0.2 M dithiothreitol, 4% w/v sodium dodecyl sulfate (SDS), 0.2% w/v bromophenol blue, 20% v/v glycerol) were heated at 95°C for 3 minutes and loaded on the gel, along with 3 µL of Unstained Protein Standard Broad Range (10-200 kDa) ladder (P7717S, NEB). Well #1 corresponds to mSandy1 (28.8 kDa), well #3 corresponds to mSandy0.7 (28.8 kDa), well #5 to mSandy2 (28.8 kDa), well #7 to mRaspberry (26.3 kDa) and well #9 to the ladder. The additional two bands around 10 kDa and 18 kDa correspond to cleaved protein products resulting from hydrolization of the N-acylimine of the chromophore upon hear denaturation. Less hydrolysis is observed for variants that mature inefficiently, due to a lower proportion of red chromophore containing RFPs in the population of molecules.

Supplementary Table S3.1. Amino-acid sequences of red fluorescent proteins.

Protein	Sequence [a]
mRojoA	MGHHHHHHGVSKGEEDNMAIIKEFMRFKTHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNANYKLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mRojo-Y197I	MGHHHHHHGVSKGEEDNMAIIKEFMRFKTHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.1	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**KLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.7	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.9	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy1	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**KLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy2	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**R**F**KLKDGGHY**V**AEVKT TYKAKKPVQLPGAY**Y**AN**I**K**M**DITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy1-L163V/L199M	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**K**M**DITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mCherry	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIK**Q**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYN**V**N**I**KLDITSHNEDYTIVEQYER**A**EGRHSTGGMDELYK

[a] Mutations from mRojoA are underlined and highlighted in bold.

3.7. References

1. Ash, C.; Dubec, M.; Donne, K.; Bashford, T., Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers in medical science* **2017,** *32* (8), 1909-1918.
2. Livet, J.; Weissman, T. A.; Kang, H.; Draft, R. W.; Lu, J.; Bennis, R. A.; Sanes, J. R.; Lichtman, J. W., Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **2007,** *450* (7166), 56-62.
3. Cai, D.; Cohen, K. B.; Luo, T.; Lichtman, J. W.; Sanes, J. R., Improved tools for the Brainbow toolbox. *Nat Methods* **2013,** *10* (6), 540-7.
4. Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **2002,** *99* (12), 7877-82.
5. Bindels, D. S.; Haarbosch, L.; van Weeren, L.; Postma, M.; Wiese, K. E.; Mastop, M.; Aumonier, S.; Gotthard, G.; Royant, A.; Hink, M. A.; Gadella, T. W., Jr., mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **2017,** *14* (1), 53-56.
6. Antaris, A. L.; Chen, H.; Diao, S.; Ma, Z.; Zhang, Z.; Zhu, S.; Wang, J.; Lozano, A. X.; Fan, Q.; Chew, L.; Zhu, M.; Cheng, K.; Hong, X.; Dai, H.; Cheng, Z., A high quantum yield molecule-protein complex fluorophore for near-infrared II imaging. *Nature communications* **2017,** *8*, 15269.
7. Megley, C. M.; Dickson, L. A.; Maddalo, S. L.; Chandler, G. J.; Zimmer, M., Photophysics and dihedral freedom of the chromophore in yellow, blue, and green fluorescent protein. *J Phys Chem B* **2009,** *113* (1), 302-8.
8. Drobizhev, M.; Molina, R. S.; Callis, P. R.; Scott, J. N.; Lambert, G. G.; Salih, A.; Shaner, N. C.; Hughes, T. E., Local Electric Field Controls Fluorescence Quantum Yield of Red and Far-Red Fluorescent Proteins. *Frontiers in molecular biosciences* **2021,** *8*, 633217.
9. Sun, Q.; Li, Z.; Lan, Z.; Pfisterer, C.; Doerr, M.; Fischer, S.; Smith, S. C.; Thiel, W., Isomerization mechanism of the HcRed fluorescent protein chromophore. *Physical chemistry chemical physics : PCCP* **2012,** *14* (32), 11413-24.
10. Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzhanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A., Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nature communications* **2020,** *11* (1), 4808.
11. Pandelieva, A. T.; Baran, M. J.; Calderini, G. F.; McCann, J. L.; Tremblay, V.; Sarvan, S.; Davey, J. A.; Couture, J. F.; Chica, R. A., Brighter Red Fluorescent Proteins by Rational Design of Triple-Decker Motif. *ACS Chem Biol* **2016,** *11* (2), 508-17.
12. Bouchiba, Y.; Cortes, J.; Schiex, T.; Barbe, S., Molecular flexibility in computational protein design: an algorithmic perspective. *Protein Eng Des Sel* **2021,** *34*.
13. Samish, I.; MacDermaid, C. M.; Perez-Aguilar, J. M.; Saven, J. G., Theoretical and computational protein design. *Annual review of physical chemistry* **2011,** *62*, 129-49.
14. Jacobs, T. M.; Williams, B.; Williams, T.; Xu, X.; Eletsky, A.; Federizon, J. F.; Szyperski, T.; Kuhlman, B., Design of structurally distinct proteins using strategies inspired by evolution. *Science* **2016,** *352* (6286), 687-90.
15. Polizzi, N. F.; DeGrado, W. F., A defined structural unit enables de novo design of small-molecule-binding proteins. *Science* **2020,** *369* (6508), 1227-1233.
16. Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L., Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc Natl Acad Sci U S A* **2010,** *107* (47), 20257-62.

17. Wannier, T. M.; Moore, M. M.; Mou, Y.; Mayo, S. L., Computational Design of the β-Sheet Surface of a Red Fluorescent Protein Allows Control of Protein Oligomerization. *PloS one* **2015,** *10* (6), e0130582.
18. Wannier, T. M.; Gillespie, S. K.; Hutchins, N.; McIsaac, R. S.; Wu, S. Y.; Shen, Y.; Campbell, R. E.; Brown, K. S.; Mayo, S. L., Monomerization of far-red fluorescent proteins. *Proc Natl Acad Sci U S A* **2018,** *115* (48), E11294-E11301.
19. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-47.
20. Lambert, T. J., FPbase: a community-editable fluorescent protein database. *Nat Methods* **2019,** *16* (4), 277-278.
21. Moore, M. M.; Oteng-Pabi, S. K.; Pandelieva, A. T.; Mayo, S. L.; Chica, R. A., Recovery of red fluorescent protein chromophore maturation deficiency through rational design. *PloS one* **2012,** *7* (12), e52463.
22. Strack, R. L.; Strongin, D. E.; Mets, L.; Glick, B. S.; Keenan, R. J., Chromophore formation in DsRed occurs by a branched pathway. *J Am Chem Soc* **2010,** *132* (24), 8496-505.
23. Kredel, S.; Oswald, F.; Nienhaus, K.; Deuschle, K.; Rocker, C.; Wolff, M.; Heilker, R.; Nienhaus, G. U.; Wiedenmann, J., mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PloS one* **2009,** *4* (2), e4391.
24. Shcherbo, D.; Murphy, C. S.; Ermakova, G. V.; Solovieva, E. A.; Chepurnykh, T. V.; Shcheglov, A. S.; Verkhusha, V. V.; Pletnev, V. Z.; Hazelwood, K. L.; Roche, P. M.; Lukyanov, S.; Zaraisky, A. G.; Davidson, M. W.; Chudakov, D. M., Far-red fluorescent tags for protein imaging in living tissues. *Biochemical Journal* **2009,** *418* (3), 567-74.
25. Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A., Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* **1999,** *17* (10), 969-73.
26. Gurskaya, N. G.; Fradkov, A. F.; Terskikh, A.; Matz, M. V.; Labas, Y. A.; Martynov, V. I.; Yanushevich, Y. G.; Lukyanov, K. A.; Lukyanov, S. A., GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett* **2001,** *507* (1), 16-20.
27. Wang, L.; Jackson, W. C.; Steinbach, P. A.; Tsien, R. Y., Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A* **2004,** *101* (48), 16745-9.
28. Shcherbakova, D. M.; Baloban, M.; Verkhusha, V. V., Near-infrared fluorescent proteins engineered from bacterial phytochromes. *Curr Opin Chem Biol* **2015,** *27*, 52-63.
29. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., Dreiding - a Generic Force-Field for Molecular Simulations. *J. Phys. Chem.* **1990,** *94* (26), 8897-8909.
30. Dunbrack, R. L., Jr., Rotamer libraries in the 21st century. *Curr Opin Struct Biol* **2002,** *12* (4), 431-40.
31. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., Iterative approach to computational enzyme design. *Proc Natl Acad Sci U S A* **2012,** *109* (10), 3790-5.
32. Dahiyat, B. I.; Mayo, S. L., Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **1997,** *94* (19), 10172-7.
33. Shortle, D.; Stites, W. E.; Meeker, A. K., Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* **1990,** *29* (35), 8033-41.
34. McCullum, E. O.; Williams, B. A.; Zhang, J.; Chaput, J. C., Random mutagenesis by error-prone PCR. *Methods Mol Biol* **2010,** *634*, 103-9.
35. Kredel, S.; Nienhaus, K.; Oswald, F.; Wolff, M.; Ivanchenko, S.; Cymer, F.; Jeromin, A.; Michel, F. J.; Spindler, K.-D.; Heilker, R.; Nienhaus, G. U.; Wiedenmann, J., Optimized and Far-Red-Emitting Variants of Fluorescent Protein eqFP611. *Chemistry & Biology* **2008,** *15* (3), 224-233.

36. Britton, H. T. S.; Robinson, R. A., CXCVIII.-Universal buffer solutions and the dissociation constant of veronal. *Journal of the Chemical Society (Resumed)* **1931,** (0), 1456-1462.
37. Winter, G., xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **2010,** *43* (1), 186-190.
38. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.
39. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 1), 72-82.
40. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007,** *40* (Pt 4), 658-674.
41. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 213-21.
42. Kantardjieff, K. A.; Rupp, B., Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **2003,** *12* (9), 1865-71.
43. Weichenberger, C. X.; Rupp, B., Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallogr D Biol Crystallogr* **2014,** *70* (Pt 6), 1579-88.
44. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 352-67.
45. Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D., electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **2009,** *65* (Pt 10), 1074-80.
46. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M., The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem* **1977,** *80* (2), 319-24.

Chapter 4: Engineering of RFP structural dynamics to rigidify the chromophore

Adam M. Damry[1,2], Sandrine Legault[1] & Roberto A. Chica[1]

[1]Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie-Curie, Ottawa, Ontario, Canada, K1N 6N5

[2]Research School of Chemistry, Australian National University, Acton, Australian Capital Territory, Australia

4.1 Statement of Contribution

This chapter is the continuation of a project started by Dr. Adam M. Damry, a former graduate student in our lab. In the work presented below, experimental design was done by Dr. Adam M. Damry and Dr. Roberto A. Chica, and the identification of clusters to engineer and all laboratory experiments were performed by Sandrine Legault.

4.2. Introduction

As we discussed up to this point in the thesis, several directed evolution and rational design approaches have been used to improve RFP brightness[3, 12-16] with some outcomes more successful than others. In addition, our work (in previous chapters) and others' have shown that the introduction of mutations in the first shell of the chromophore pocket often leads to unwanted blue-shifted fluorescence[14-16]. This may be due to changes in the electrostatic environment of the chromophore resulting in perturbation of the delocalized electrons of the RFP chromophore, thereby increasing the energy between the excited and ground singlet states' energy levels (described in Chapter 1). One way to avoid this problem would be to introduce beneficial mutations distal from the first shell of the

chromophore. However, distal mutations that result in enhanced brightness are difficult to predict and are therefore usually identified by directed evolution[3] which is expensive in terms of resources and effort. A more rational method capable of pin-pointing distal mutagenesis hotspots in the RFP structure to increase brightness would be highly beneficial.

Such an approach has been proposed by Damry *et al.* (2019) who performed protein nuclear magnetic resonance (NMR) spectroscopy on a set of mCherry-related RFPs with various quantum yields and brightness (**Figure 4.1**) to compare structural dynamics on a residue-by-residue basis[1]. Peaks obtained from NMR ^{1}H-^{15}N HSQC spectra correspond to the backbone amide of a residue and can provide information on the dynamics of the protein backbone. More precisely, peak intensities provide information on local backbone dynamics fluctuations in the microsecond-to-millisecond or nanosecond timescales[17, 18]. In this analysis, deviation of peak intensities from average values were obtained for the set of RFPs. This analysis identified 25 positive correlations between residue rigidity and brightness, where peak intensity deviation was smallest in RFPs with increasing brightness (**Supplementary Figure S4.1**).

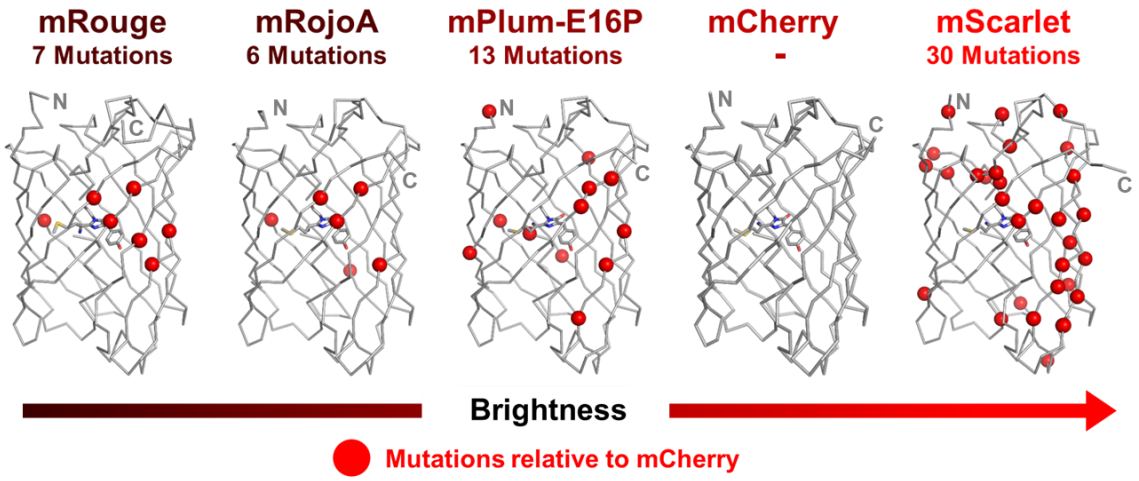


Figure 4.1. Selected mCherry-related RFPs containing different numbers of mutations and levels of brightness. A total of five related RFPs with increasing brightness were chosen for protein NMR experiments, shown as gray ribbons: mRouge (PDB ID: 3NED[19]), mRojoA (PDB ID: 3NEZ[19]), mPlum-E16P (PDB ID: 4H3L[20]), mCherry (PDB ID: 2H5Q[21]), and mScarlet (PDB ID: 5LK4[3]). The chromophore is shown as gray sticks and mutated positions relative to the protein of reference, mCherry, are shown as red spheres. Additional mutations located at the ends of the N- and C-termini are not shown. Figure adapted from Damry *et al.* (2019)[1].

In a similar fashion, relaxation measurements, which determine the rate of decay of the NMR signal, were performed to obtain additional information on backbone dynamics that occur on the picosecond to nanosecond timescales[22]. The relaxation rate is mainly influenced by the tumbling rate of the protein in solution directly reflected by the correlation time value, which is defined as the average time for a molecule to rotate by one radian. A shorter correlation time indicates rapid tumbling and a longer correlation time indicates slower tumbling. When local motions occur at a rate that is faster than the protein tumbling rate, this creates an "apparent tumbling rate" reflected by an "apparent correlation time" value. Therefore, variations in apparent correlation time values between residues represent local backbone fluctuations. Based on the extrapolated values, longer correlation times were obtained for 11 positions in RFPs with increasing brightness, thus were

identified as having a positive correlation between residue rigidity and RFP brightness (**Supplementary Figure S4.1**). Three of these positions were also identified as less mobile in the HSQC spectrum. Hence, a total of 33 positions were identified throughout the RFP scaffold where residue rigidity positively correlated to brightness (**Figure 4.2**).

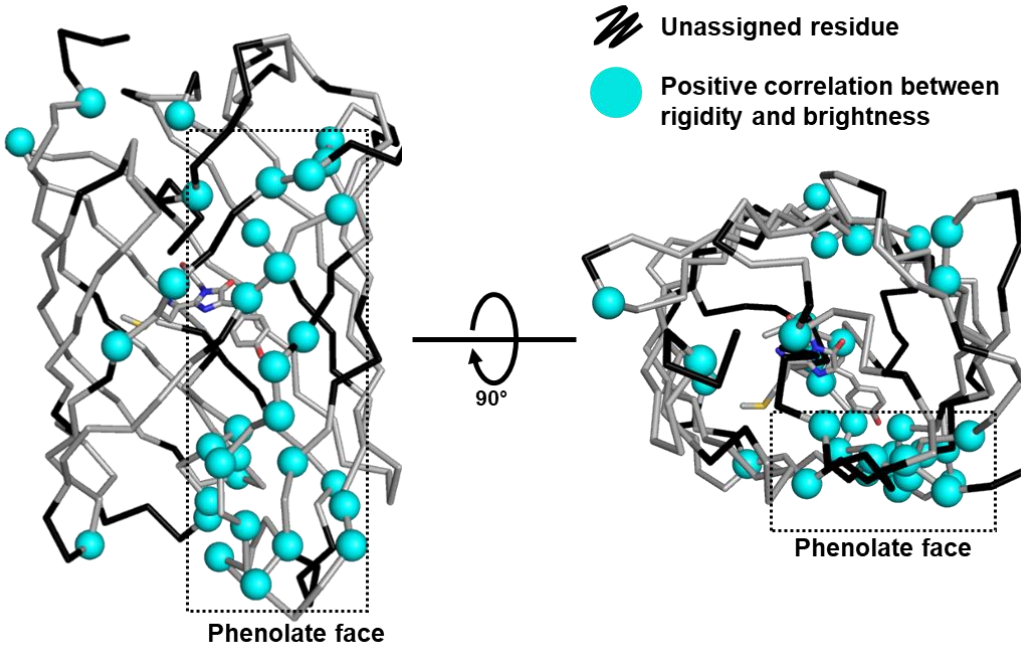


Figure 4.2. Clustering of positions where residue rigidity positively correlates to brightness. The 33 positions where positive correlations were observed are shown as cyan spheres on the mCherry backbone represented by a gray ribbon, and the chromophore is shown as gray sticks. Positions where correlations could not be assessed (because they could not be unambiguously identified by NMR either in both mRojoA and mRouge or in one other or more of the selected RFPs) are shown as black portions of the ribbon. Clustering of positive correlations occurs on the phenolate face of the RFP barrel (outlined by a dotted line).

Based the NMR analysis results, we observed clusters of adjacent residue positions in 3D space showing a similar correlation. Herein, we hypothesize that protein regions where residue dynamics were found to positively correlate with quantum yield can be engineered to increase brightness in related RFPs. We also propose that surface residues

may be potential engineering targets to identify beneficial distal mutations which possess the ability to mimic the stabilizing effect of RFP dimerization that has previously been shown to increase RFP brightness[23, 24], without affecting the electronic chromophore environment. Through stabilizing side chain interactions, surface mutations could rigidify the dimerization interface of the β-barrel, on which they are located, and thus indirectly rigidify the chromophore environment to increase quantum yield.

To test our hypothesis, we performed combinatorial site-saturation mutagenesis on two clusters of residues on the surface of the mCherry β-barrel (**Figure 4.3** methodology workflow), each comprising three residue positions (194, 196, 220 and 144, 145, 198). These separate combinatorial libraries of 20^3 variants were then screened for enhanced brightness via fluorescence-activated cell sorting (FACS). Our screening results did not yield significantly brighter mCherry variants, despite displaying enhanced fluorescence intensity during screening. Thus, our results remain inconclusive towards our hypothesis that engineering clusters of surface residues where positive correlations between residue rigidity and RFP brightness were observed by NMR analysis can increase RFP brightness. We suggest that core residues or a more direct screening method such as fluorescence lifetime measurements should be considered in future work.

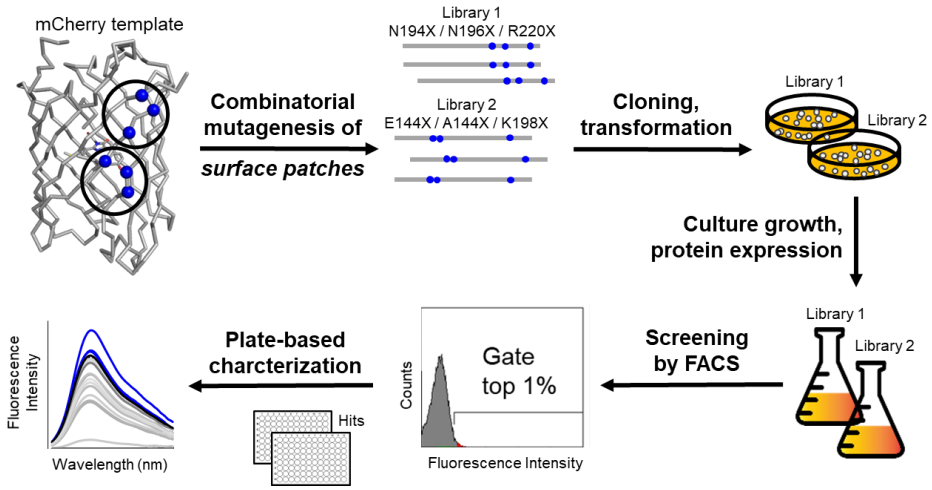


Figure 4.3. Methodology workflow. Based on the clustering information obtained by NMR, our goal is to increase the brightness of mCherry (PDB ID: 2H5Q) using combinatorial site-saturation mutagenesis of two rigidity-correlated surface patches (N194X / N196X / R220X and E144X / A145X / K198X) followed by high-throughput screening for enhanced brightness by FACS and characterization of the best mutants identified.

4.3. Results

Previous work, rationale and engineering hypothesis

It was previously proposed by Damry *et al.* (2019) that protein NMR spectroscopy, a technique used to obtain information about protein structure and dynamics, could be used to obtain correlations between residue dynamics and RFP brightness, identifying engineering hotspots to create brighter RFP variants[1]. In order to identify these correlations, Damry *et al.* (2019) acquired ^{1}H-^{15}N HSQC spectra and measured correlation times on a set of five related RFPs containing between 6 and 30 mutations relative to the widely used RFP mCherry: namely, mRouge, mRojoA, mPlum-E16P, mCherry and mScarlet (**Figure 4.1**, **Table 4.1**). Importantly, these RFPs span a wide range of quantum yields (0.02 – 0.76) and brightness values (1 – 67 mM^{-1} cm^{-1}) (**Table 4.1**). This analysis revealed 33 positions where residue rigidity positively correlated to quantum yield and

brightness. When these positions were mapped onto the mCherry RFP structure, they were found to cluster on the face of the β-barrel that is adjacent to the phenolate moiety of the chromophore, which we will refer to as the "phenolate face" (**Figure 4.2**). This phenolate face is also known as an RFP dimerization interface[24].

Table 4.1. Spectroscopic characteristics of selected mCherry-related RFPs spanning a wide range of quantum yields and brightness. RFPs in this table are shown in decreasing order of brightness, where mScarlet is the brightest and mRouge is the dimmest.

Protein	λ_{ex} (nm)	λ_{em} (nm)	Φ_F	ε ($mM^{-1}cm^{-1}$)	Brightness ($mM^{-1}cm^{-1}$)	# Mutations relative to mCherry
mScarlet	569 ± 1	592 ± 1	0.75 ± 0.04	93 ± 14	70	30
mCherry	586 ± 2	609 ± 1	0.23 ± 0.01	87 ± 3	20	-
mPlum-E16P	590 ± 2	621 ± 1	0.13 ± 0.01	82 ± 3	11	13
mRojoA	596 ± 2	632 ± 2	0.023 ± 0.001	68 ± 9	2	6
mRouge	604 ± 3	633 ± 3	0.018 ± 0.001	64 ± 5	1	7

λ_{ex} and λ_{em} correspond to excitation and emission maxima.
The brightness is the product of the quantum yield (Φ_F) and the extinction coefficient (ε).
$n \geq 3$ independent experiments in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4). Each independent experiment consisted of triplicate measurements performed on a single protein batch.

This dimerization interface has been studied in previous work by Alford *et al.* (2012) who showed that heterodimerization of a dim monomeric DsRed-derived RFP with another subunit, in which key residues involved in chromophore maturation were mutated to convert it into a non-fluorescent protein, resulted in a 2.8-fold enhanced quantum yield and 10-fold enhanced brightness with no significant impact on the emission wavelength[23]. The authors suggested that the increased brightness may result from an indirect rigidification of the chromophore environment in a way that stabilized the chromophore in a coplanar conformation thus reducing its loss of energy via non-radiative decay. The benefit of RFP dimerization on brightness has also been demonstrated by Campbell *et al.*

(2002) by which generation of monomeric RFP from its tetrameric precursor DsRed (Φ_F = 0.68) yielded an approximately 3-fold decrease in quantum yield (mRFP1, Φ_F = 0.25), in contrast with their dimeric mutant (dimer2, Φ_F = 0.69) which exhibited a similar quantum yield value relative to the tetrameric parent protein[24].

Based on these findings, we hypothesize that the brightness of mCherry can be enhanced by targeting residues located on the surface of the protein phenolate face and whose side chains have the ability to interact and potentially form stabilizing interactions that mimic the effect of RFP dimerization and indirectly rigidify the RFP chromophore in a coplanar conformation. Engineering brighter mCherry variants with no undesired blue-shifted emission maximum could provide useful RFPs. In order to do this, we chose to mutate two different patches of residues located on the mCherry β-barrel for which similar correlations based on the NMR study by Damry *et al.* (2019) were observed (**Figure 4.4a**). For each surface patch, the possibility of side chain interaction was defined as side chains pointing outside of the RFP barrel with potential rotamers that point in converging directions, and having a C_α- C_α distance cut-off of 8 Å, previously shown to be useful to identify residue-residue contacts[25], between one residue and at least one other residue in that cluster (**Figure 4.4b**). Because it is difficult to predict how surface residue side chains may interact and which combination of these residues would be able to mimic dimerization, we used combinatorial site-saturation mutagenesis yielding two surface patch libraries (20^3 variants each) , N194X / N196X / R220X and E144X / A145X / K198X (**Figure 4.4b**). In the case of the latter library, K198 was not assigned by NMR but is adjacent to several positively correlated residue positions. Because the NMR ^{1}H-^{15}N HSQC spectrum and the

apparent correlation time provide information on a residue and the ones strictly around it, it is acceptable to include K198 in our mutagenesis targets.

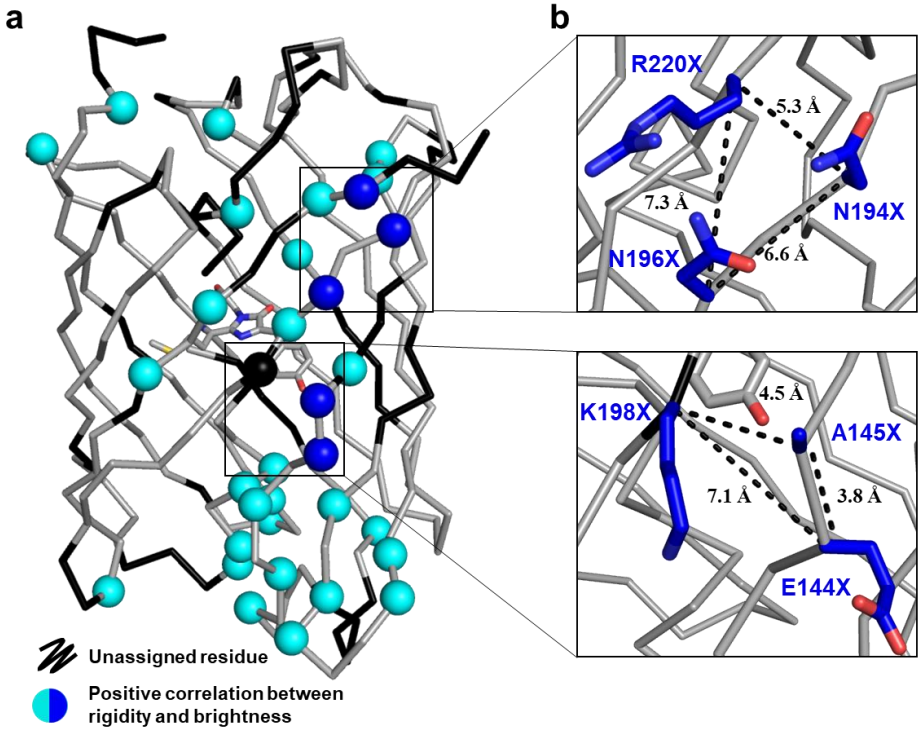


Figure 4.4. Surface patches chosen for combinatorial mutagenesis in mCherry. (a) Mapping of chosen surface patches (outlined) onto the phenolate face of mCherry (PDB ID: 2H5Q, shown as a ribbon), clustering positions where positive rigidity-brightness correlations was observed (shown as cyan or blue spheres). Positions where correlations could not be unambiguously identified by NMR are shown as black portions of the ribbon. (b) Two surface patches chosen for combinatorial site-directed mutagenesis, defined as surface residues (shown as blue sticks) whose C_α- C_α distances are less than 8 Å (measured using PyMOL and shown as dashed lines), and whose potential mutant rotamers may have the ability to interact as speculated by visual inspection.

Combinatorial site-saturation mutagenesis

Using mCherry as the starting template, we performed combinatorial site-saturation mutagenesis at surface residue positions 194, 196, 220 and 144, 145, 198, separately, in order to obtain two distinct libraries of mCherry surface patch mutants. Libraries were

generated using overlap-extension polymerase chain reaction (PCR) (**Supplementary Figure S4.2**) with respective primers containing degenerate NNS codons, where N corresponds to any nucleotide (A, T, G and C) and S corresponds to G or C, thereby encoding all 20 canonical amino acids. Sequencing results revealed that the libraries contained the desired mixtures of nucleotides at each designed positions (**Supplementary Figure S4.3**), thus allowing full coverage of residue combinations for each surface patch library to be tested experimentally.

High-throughput screening

Full coverage of each library corresponds to 20^3 possible protein sequences, emphasizing the need for a high-throughput pre-screening method. In order to screen for enhanced brightness, we submitted the cells to FACS using an excitation laser (561 nm) and emission detector (614 nm) adapted for mCherry wavelengths. The top 1-10% of cells of each library were sorted, gated based on highest fluorescence intensity (**Supplementary Figure S4.4**). Two additional rounds of cell sorting were performed on the cells collected to reduce the amount of false positives and enrich the true positives in the sorted populations (**Supplementary Figure S4.5**). After the three rounds, FACS analysis demonstrated that the two sorted populations of the surface patch libraries were globally brighter than mCherry, as evidenced by the shifts in the population peaks of the mutants relative to that of the mCherry control (**Figure 4.5**). During the third round of FACS, the top 1% of the cells of each library were sorted into 96 well plates for plate-based characterization.

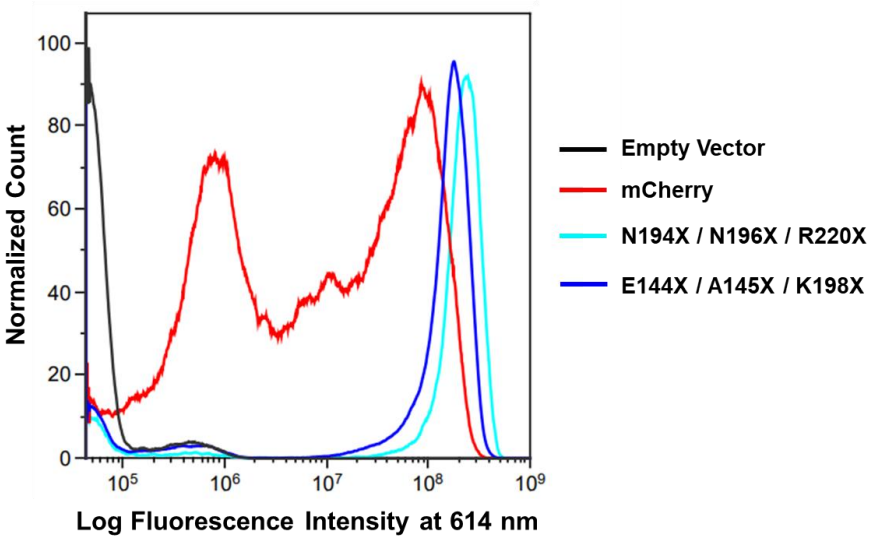


Figure 4.5. Brighter surface patch mutant populations after three rounds of FACS. After three rounds of FACS screening for enhanced fluorescence intensity (λ_{ex} = 561 nm, λ_{em} = 614 nm) using a MoFlo Astrios EQ cell sorter (Beckman Coulter), the surface patch mutant libraries in *E. coli* BL21-Gold (DE3) cells were enriched for brighter populations. This is shown by the shift of the population peaks of the N194X/N196X/R220X and E144X/A145X/K198X variants, (shown as cyan and blue curves, respectively) relative to that of the mCherry population (shown as a red curve).

Spectroscopic characterization

Sequencing results of the best mutants revealed that two originated from the surface patch library 1, mCherry-N196A (mCherryA) and mCherry-N194Q/N196L/R220V (mCherryQLV), and one from the surface patch library 2, mCherry-A145I/K198G (mCherryIG). All three variants displayed similar spectral properties to their parent protein mCherry, with efficient red chromophore maturation as evidenced by the lack of green chromophore (λ = 510 nm) in their absorption spectra. The mutants also displayed similar emission wavelength maxima to mCherry (**Figure 4.6**).

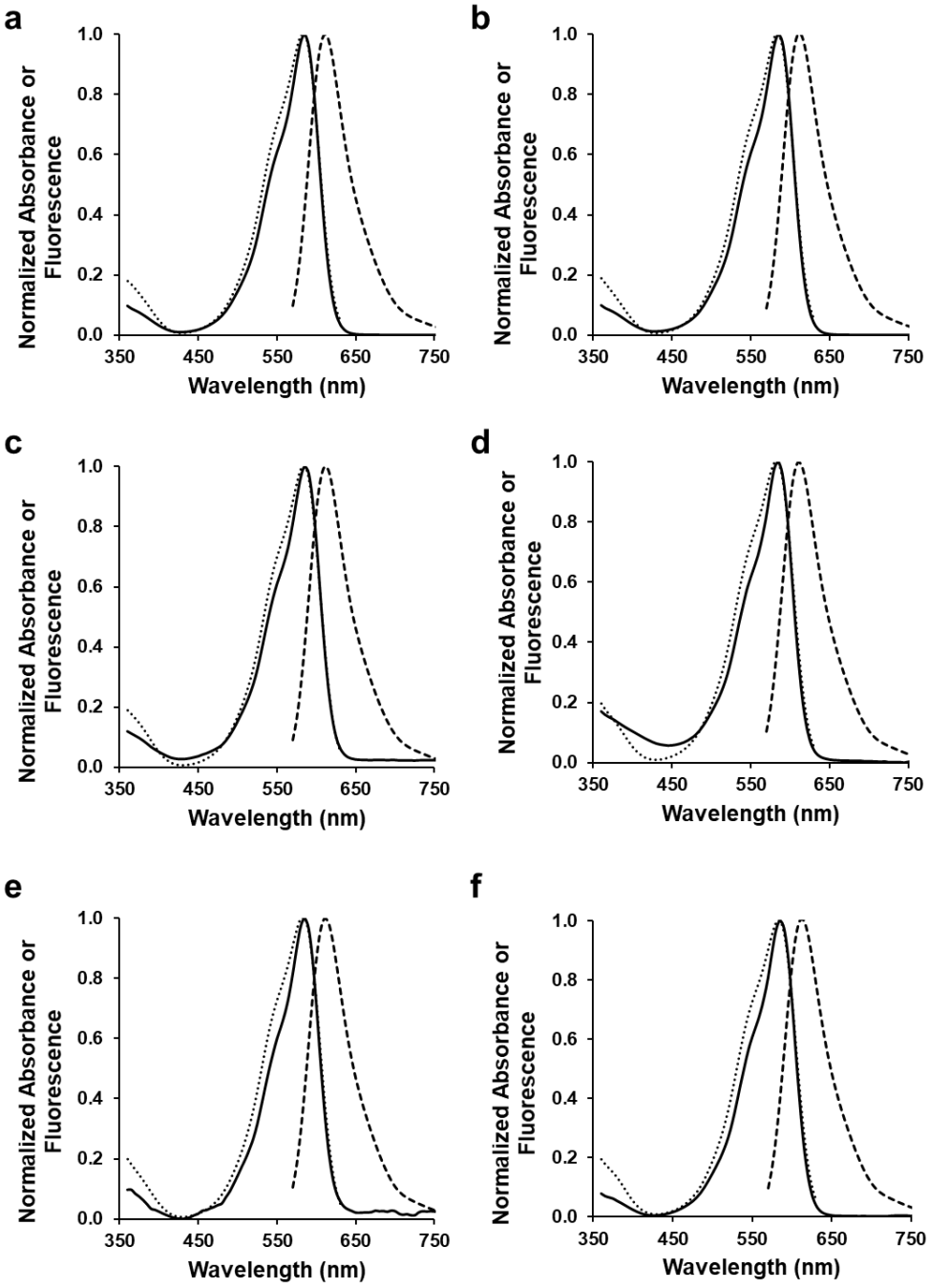


Figure 4.6. Fluorescence and absorption spectra of mCherry surface patch mutants. Absorption (solid lines), excitation (dotted lines) and emission (dashed lines) spectra of purified (a) mCherry, (b) mCherryA, (c) mCherryQLV, (d) mCherryIG, (e) mCherryIAG and (f) mCherryIQLGV in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4). Spectra were measured at room temperate using an Infinite M1000 (Tecan) or Spectramax i3 (Molecular Devices) plate reader.

However, mCherryA, mCherryQLV and mCherryIG displayed quantum yields of 0.24, 0.24 and 0.23, respectively, which are not significantly higher than that of mCherry (Φ_F = 0.23 ± 0.01) (**Table 4.2**). We then aimed to investigate whether combining the different mutant surface patches would result in an increased quantum yield via synergistic effect of the mutations. In order to do so, we characterized the following mutants: mCherry-A145I/N196A/K198G (mCherryIAG) and mCherry-A145I/N194Q/N196L/K198G/R220V (mCherryIQLGV). These variants displayed quantum yields of 0.21 and 0.22 with no significant increase in brightness (**Table 4.2**). However, these observations remain inconclusive towards our hypothesis since only a subset of positions were evaluated, all of which were restricted to surface residues.

Table 4.2. Spectral properties of the best mCherry surface patch mutants.

Protein	λ_{abs} (nm)	λ_{ex} (nm)	λ_{em} (nm)	Quantum Yield	Extinction Coefficient ($mM^{-1}cm^{-1}$)	Brightness ($mM^{-1}cm^{-1}$)	Mutations Relative to mCherry					
							144	145	194	196	198	220
mCherry	586 ± 1	586 ± 2	609 ± 1	0.23 ± 0.01	87 ± 3	20	E	A	N	N	K	R
Surface patch library 1												
mCherryA	585	585	610	0.24	-	-	-	-	-	A	-	-
mCherryQLV	585	585	610	0.24	-	-	-	-	Q	L	-	V
Surface patch library 2												
mCherryIG	585	585	613	0.23	-	-	-	I	-	-	G	-
Combined mutations												
mCherryIAG	590	585	610	0.21	91	19	-	I	-	A	G	-
mCherryIQLGV	590	585	610	0.22	98	22	-	I	Q	L	G	V

λ_{abs}, λ_{ex} and λ_{em} correspond to the absorption, excitation, and emission maxima, respectively.
Brightness is the product of the quantum yield and the extinction coefficient.
Experiments for mCherry were performed using n = 3 independent biological replicates (mean ± sd). n = 1 independent experiment for all other variants. Each independent experiment consisted of triplicate measurements performed on a single protein batch.
All values are at pH 7.4 in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate).

4.4. Discussion

In this work, we used combinatorial site-saturation mutagenesis at surface positions where positive correlations between residue rigidity and RFP brightness were identified by protein NMR spectroscopy. We speculated that mutating these selected surface patches located on a known dimerization interface would mimic dimerization-dependent enhanced fluorescence and therefore increase RFP brightness. Our spectroscopic characterization results demonstrated that the mutants did not display undesired hypsochromic shifts, supporting the fact that introducing mutations distal from the chromophore pocket does not perturb the electrostatic environment of the chromophore. However, our results also showed that none of the mutants exhibited significant increases in quantum yield or overall brightness, despite displaying enhanced fluorescence intensity by FACS. It is important to note that the fluorescence intensity observed can be influenced by improved maturation, solubility, folding or higher protein concentration. Based on our results, we speculate that the enhanced fluorescence intensity observed during high-throughput screening may result from improved solubility, albeit further experiments would be required to support this. In fact, most engineering efforts focused on the redesign of protein surfaces have been made to exclusively enhance protein solubility[26-28]. A more direct screening methodology should be considered to avoid discrepancy between results from whole-cells and purified samples. For example, high-throughput fluorescence lifetime imaging could be considered since fluorescence lifetime is directly proportional to quantum yield.

In addition, it may not be possible to rigidify regions of the RFP dimerization interface by solely mutating surface residues. This could be partly explained by the effect of protein solvation whereby the strength of electrostatic interactions is weakened in the

presence of solvent[29]. Therefore, surface residues whose side chains have the ability to interact via salt bridges or hydrogen bonds would be destabilized due to interactions with the solvent. Thus, local packing and rigidity would not be enhanced. In the case of aliphatic residues, it has recently been shown that introducing hydrophobic residues in surface loops can stabilize protein folds by improving local packing and solubility[30]. However, this could be limited when applied to hydrophobic surface residues located on β-sheet secondary structures which have fewer degrees of freedom for the protein fold to remain stable. In addition, such hydrophobic surface patches may destabilize the protein fold due to reduced entropy when they are exposed to solvent[31]. Nevertheless, in both cases, it would be worth investigating other residue clusters such as those that are oriented into the protein core in which local packing could be more efficiently achieved due to lack of solvent access.

4.5. Conclusions and Perspectives

Taken altogether, our results demonstrated that engineering of the proposed surface patches did not lead to improved RFP brightness. Although our results in Chapter 3 suggest that it is possible to identify surface mutations that contribute to increased quantum yield as observed in mSandy2, the engineering targets in this chapter did not provide beneficial combinations of mutations. Therefore, we propose that future work should investigate other residue positions distal from the chromophore environment, such as those whose side chains point towards the inside of the RFP barrel. Combinatorial mutant clusters of buried residues, for example clusters V22X / 56X / I60X and A57X / D59X / T140X / L165X (**Figure 4.7**), may have greater potential to improve local packing and decrease local structural dynamics to indirectly rigidify the chromophore pocket and increase RFP

quantum yield and brightness. Future experiments of the proposed clusters could be combined with the computation aliphatic packing approach described in Chapter 2 and could also benefit from a more direct screening method such as high-throughput lifetime imaging.

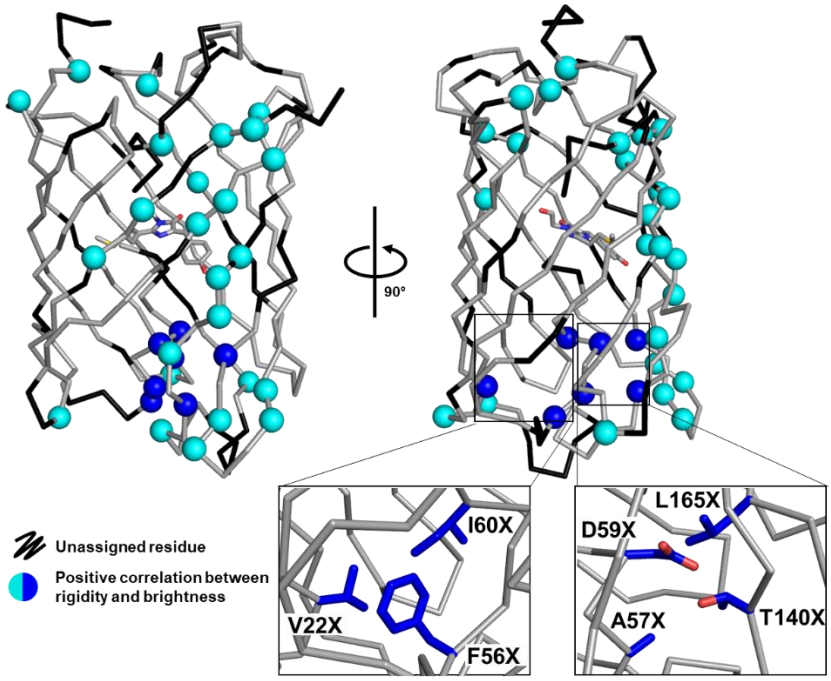
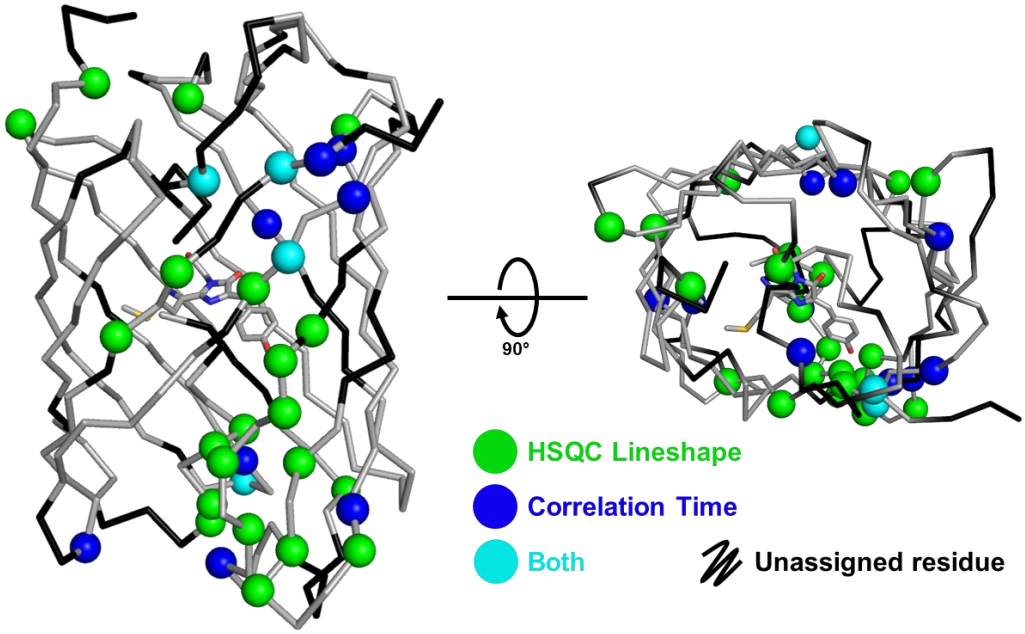


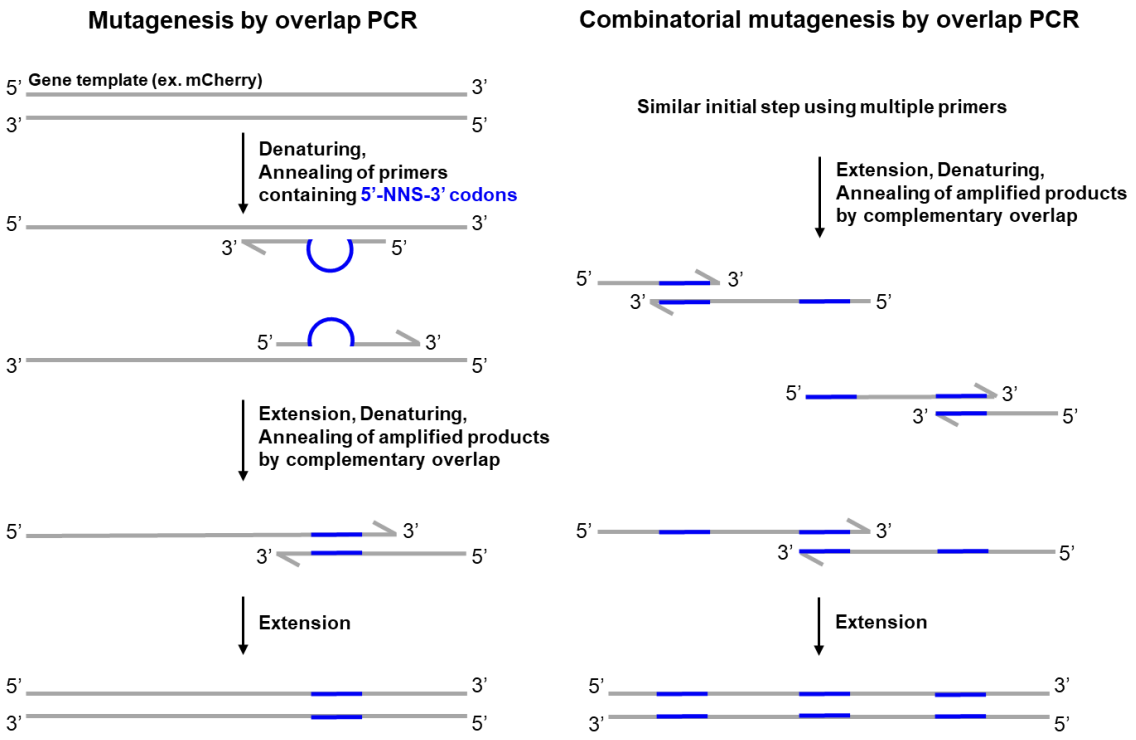
Figure 4.7. Potential core clusters to target to engineer RFP brightness. Residue positions where positive rigidity-brightness correlations were observed are shown as cyan or blue spheres. Positions where correlations could not be unambiguously identified by NMR are shown as black portions of the ribbon. Future work should focus on combinatorial site-saturation mutagenesis at positively correlated residue positions located inside the protein core (shown as blue sticks): V22X / 56X / I60X and A57X / D59X / T140X / L165X.

4.6. Supplementary Information

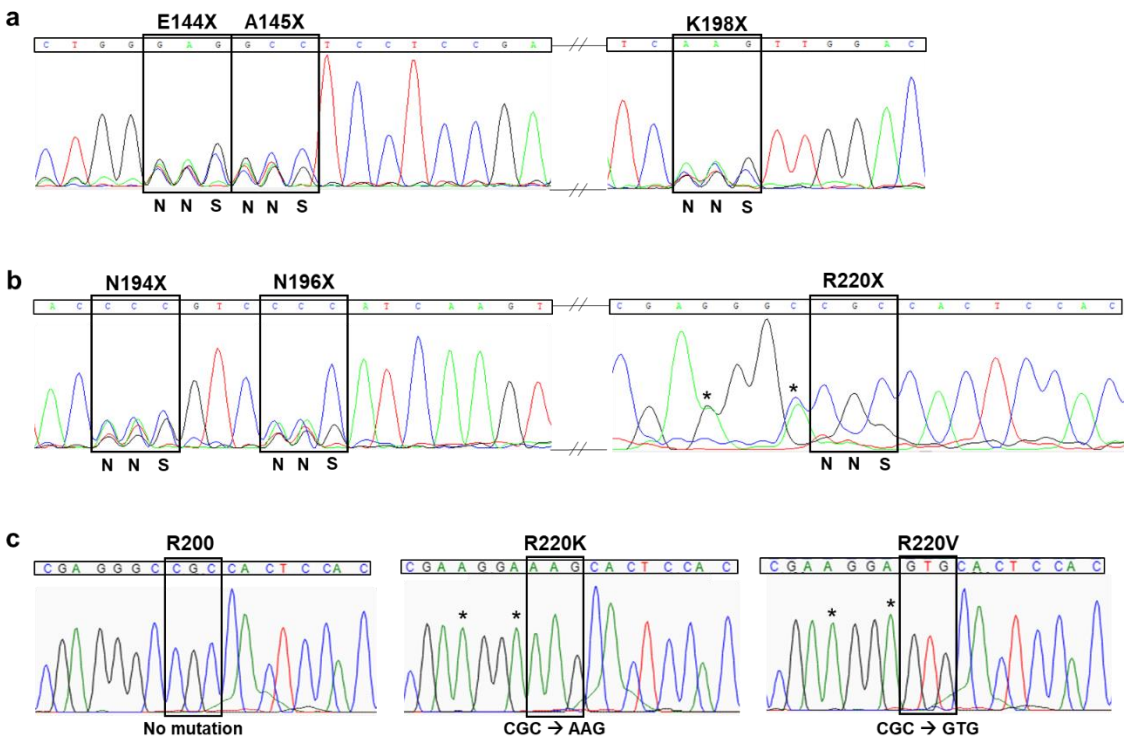
Supplementary Information contains four figures.



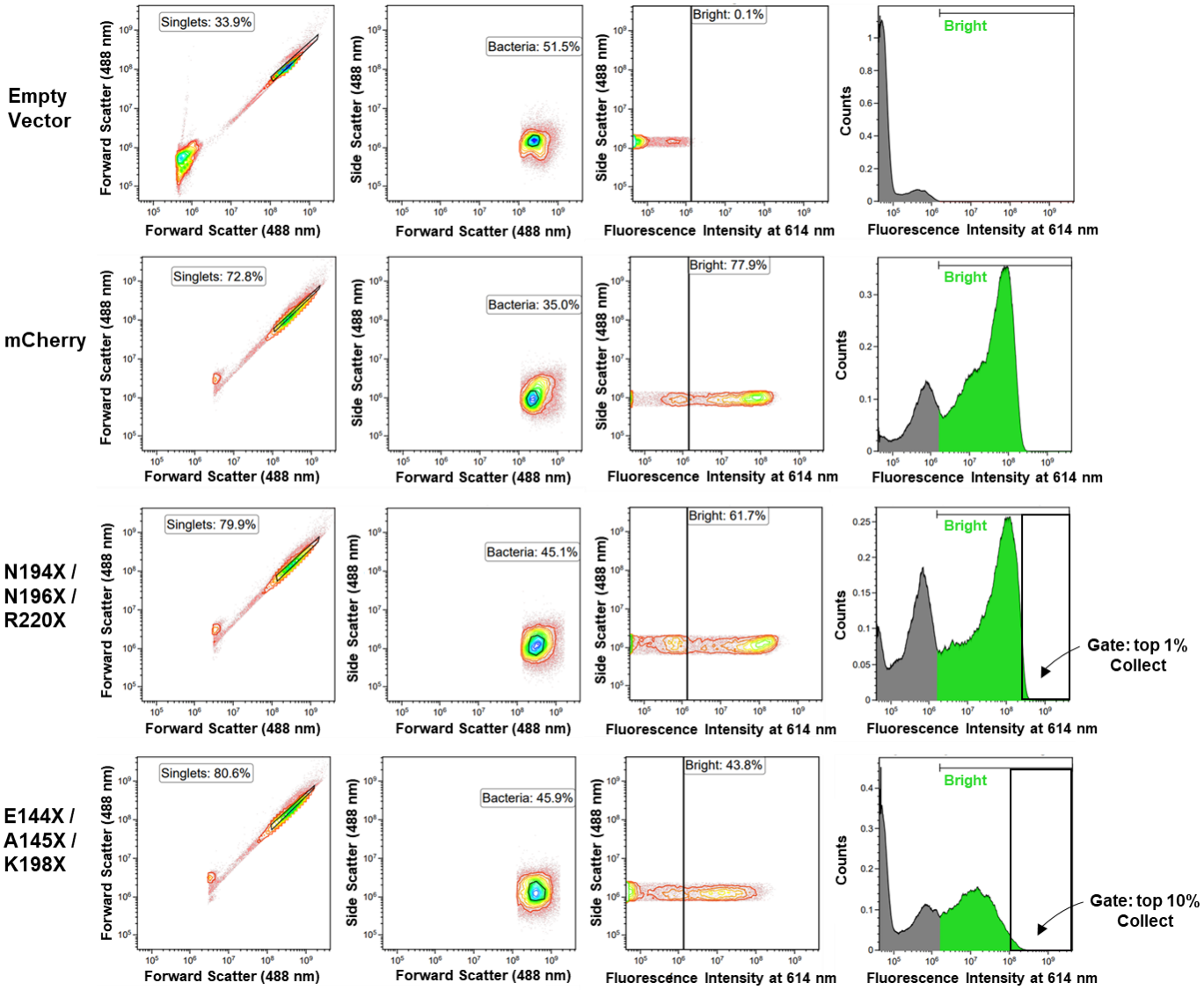
Supplementary Figure S4.1. Positive correlations between residue rigidity and RFP quantum yield and brightness using different NMR measurements. The 33 positions where positive correlations were observed are shown as spheres on the mCherry backbone represented by a gray ribbon, and the chromophore is shown as gray sticks. Twenty-five positions were identified by HSQC analysis (green) and 11 positions were identified by correlation time measurements (blue), 3 of these 33 positions were identified using both NMR analyses (cyan). Positions where correlations could not be assessed (because they could not be unambiguously identified by NMR either in both mRojoA and mRouge or in one other or more of the selected RFPs) are shown as black portions of the ribbon. Of note, the E144X/A145X/K198 surface patch corresponds to residues whose correlations were identified by HSQC. In contrast, the N196X/N194X/R220X surface patch corresponds to residues whose correlations were identified by correlation time measurements, with N196X identified using both methods. Figure adapted from Damry *et al.* (2019)[1].



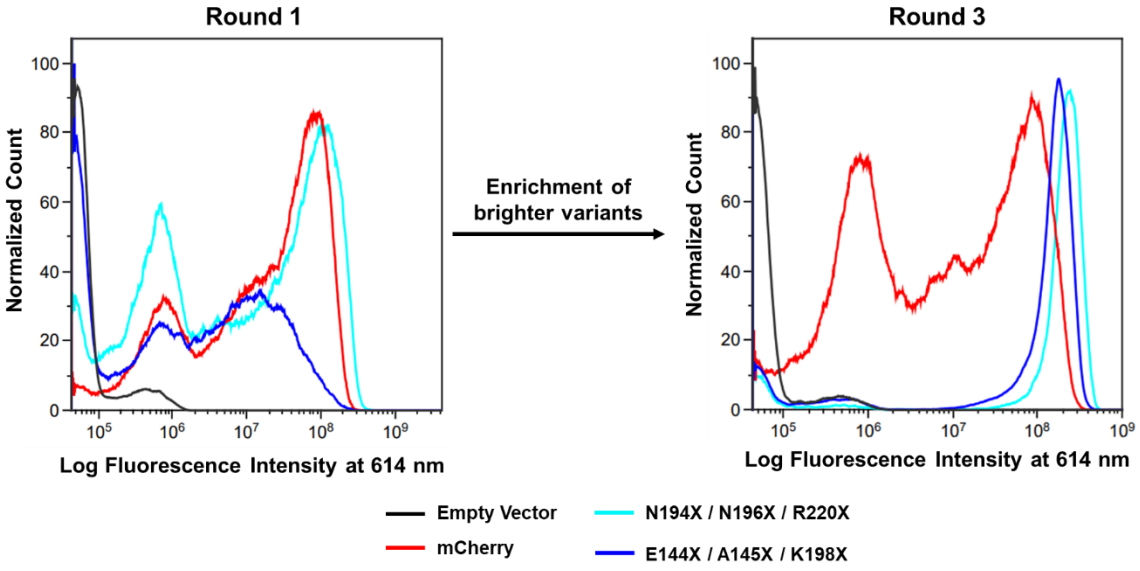
Supplementary Figure S4.2. General overview of combinatorial site-saturation mutagenesis using overlap-extension polymerase chain reaction (PCR). Surface patch libraries (N194X / N196X / R220X and E144X / A145X / K198X) were generated by combinatorial site-saturation mutagenesis on the mCherry gene template using complementary primers containing degenerate NNS codons for each position (see **5.1. Materials and Methods**).



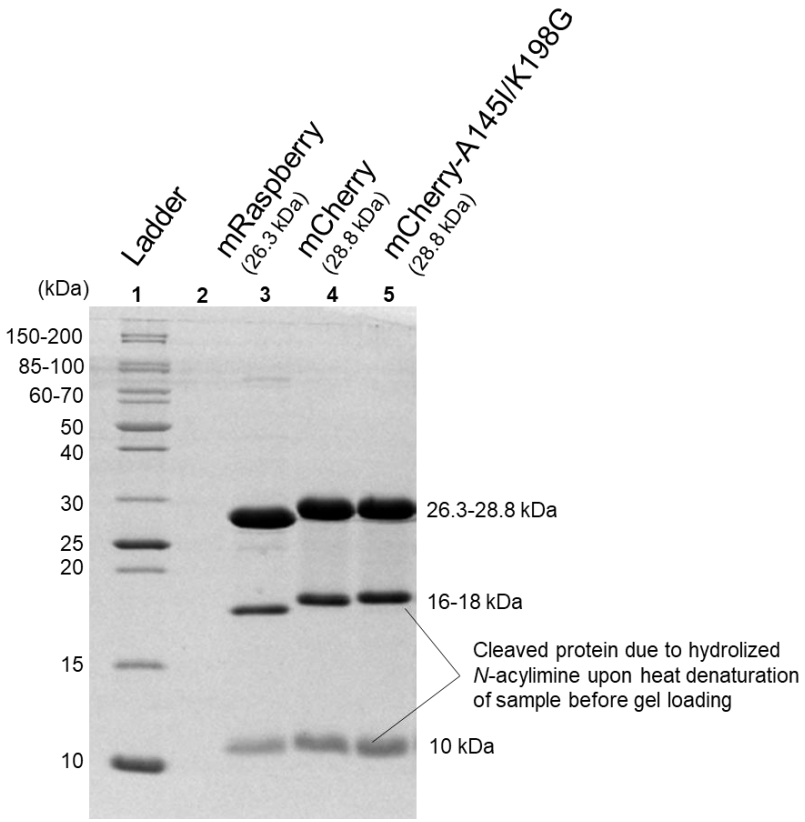
Supplementary Figure S4.3. DNA sequencing results of surface patch libraries. Plasmid DNA of each surface patch library (N194X/N196X/R220X and E144X/A145X/K198X) was sent for sequencing at the Ottawa Hospital Research Institute and Genome Quebec. a) Sequencing results of the E144X / A145X / K198X surface patch library, showing incorporation of NNS codons at desired sites. b) Sequencing results of the N194X/N196X/R220X surface patch library showing incorporation of NNS codons at desired sites. N corresponds to any nucleotide, A (green), T (red), C (blue) or G (black), and S corresponds to C (blue) or G (black). Asterisks represent silent mutations incorporated into the designed primers to reduce the G/C content and therefore reduce the primer melting temperature to one that is low enough to permit proper annealing ($T_{annealing}$ < 72°C). c) Sequencing results of individual N194X/N196X/R220X library mutants. Because sequencing results of R220X gave nucleotide signals close to the baseline (which we assumed was due to that codon being close to the end of the sequencing template thus less accurately sequenced), we isolated individual colonies and sent them for sequencing. Results show the incorporation of the silent mutations expected and variety in codon sequence as expected from "NNS" usage.



Supplementary Figure S4.4. Gated mCherry surface patch libraries for cell sorting after the first round of FACS. Parameters used to select stringent gates for cell sorting of the mCherry surface patch N194X/N196X/R220X and E144X/A145X/K198 libraries in BL21-Gold DE3 cells. This figure provides a summary of the FACS data acquired during the first round. All axes are logarithmic. Forward and side scatters reflect the size and granularity of the bacterial cells, respectively. In the first plots, the cells are gated so that only singular non-aggregated cells are selected. In the second plots, the cells are gated according to their shape so that dying cells or odd particles are excluded. The third plots are gated to represent the fluorescent cells (named "bright") based on what is observed in the empty vector (negative control). These fluorescent populations are represented as green areas on the histograms (on the right). The top 1% and the top 10% of these bright populations of the N194X/196X/R220X and the E144X/A145X/K198X libraries were collected, respectively.



Supplementary Figure S4.5. Enrichment of brighter mCherry surface patch mutants. During the first round of cell sorting, the N194X/N196X/R220X library (shown as a cyan curve) seemed to possess variants with higher fluorescence relative to mCherry (shown as a red curve). Whereas the E144X/A145X/K198X library (shown as a blue curve) displayed a dimmer RFP population. This could be due to the fact that the positions mutated in that surface patch are directly next to the chromophore phenolate, despite being surface residues. Therefore, slight destabilization could hinder chromophore maturation and fluorescence emission.



Supplementary Figure S4.6. Representative SDS-PAGE of purified surface patch mutant mCherry-A145I/K198G (mCherryIG), and mCherry and mRaspberry controls. Purified proteins in 1X SDS loading dye (0.1 M Tris-HCl pH 6.8, 0.2 M dithiothreitol, 4% w/v sodium dodecyl sulfate (SDS), 0.2% w/v bromophenol blue, 20% v/v glycerol) were heated at 95°C for 3 minutes and loaded on the gel, along with 3 µL of Unstained Protein Standard Broad Range (10-200 kDa) ladder (P7717S, NEB). Well #1 corresponds to the ladder, well #3 corresponds to mRaspberry (26.3 kDa), well #4 to mCherry (28.8 kDa) and well #5 to mCherryIG (28.8 kDa). The additional two bands around 10 kDa and 18 kDa correspond to cleaved protein products resulting from hydrolization of the N-acylimine of the chromophore upon hear denaturation.

4.7. References

1. Damry, A. M. From protein sequence to motion to function: Towards the rational design of functional protein dynamics. Doctoral dissertation, Université d'Ottawa/University of Ottawa, 2019.
2. Ransom, E. M.; Ellermeier, C. D.; Weiss, D. S., Use of mCherry Red fluorescent protein for studies of protein localization and gene expression in Clostridium difficile. *Appl Environ Microbiol* **2015,** *81* (5), 1652-60.
3. Bindels, D. S.; Haarbosch, L.; van Weeren, L.; Postma, M.; Wiese, K. E.; Mastop, M.; Aumonier, S.; Gotthard, G.; Royant, A.; Hink, M. A.; Gadella, T. W., mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **2017,** *14* (1), 53-56.
4. Nasu, Y.; Shen, Y.; Kramer, L.; Campbell, R. E., Structure- and mechanism-guided design of single fluorescent protein-based biosensors. *Nat Chem Biol* **2021,** *17* (5), 509-518.
5. Qian, Y.; Cosio, D. M. O.; Piatkevich, K. D.; Aufmkolk, S.; Su, W. C.; Celiker, O. T.; Schohl, A.; Murdock, M. H.; Aggarwal, A.; Chang, Y. F.; Wiseman, P. W.; Ruthazer, E. S.; Boyden, E. S.; Campbell, R. E., Improved genetically encoded near-infrared fluorescent calcium ion indicators for in vivo imaging. *PLoS Biol* **2020,** *18* (11), e3000965.
6. Bischof, H.; Burgstaller, S.; Waldeck-Weiermair, M.; Rauter, T.; Schinagl, M.; Ramadani-Muja, J.; Graier, W. F.; Malli, R., Live-Cell Imaging of Physiologically Relevant Metal Ions Using Genetically Encoded FRET-Based Probes. *Cells* **2019,** *8* (5).
7. Livet, J.; Weissman, T. A.; Kang, H.; Draft, R. W.; Lu, J.; Bennis, R. A.; Sanes, J. R.; Lichtman, J. W., Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **2007,** *450* (7166), 56-62.
8. Clement, G. T.; Hynynen, K., Ultrasound phase-contrast transmission imaging of localized thermal variation and the identification of fat/tissue boundaries. *Phys Med Biol* **2005,** *50* (7), 1585-600.
9. Winnard, P. T.; Kluth, J. B.; Raman, V., Noninvasive optical tracking of red fluorescent protein-expressing cancer cells in a model of metastatic breast cancer. *Neoplasia* **2006,** *8* (10), 796-806.
10. Ash, C.; Dubec, M.; Donne, K.; Bashford, T., Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers Med Sci* **2017,** *32* (8), 1909-1918.
11. Drobizhev, M.; Molina, R. S.; Callis, P. R.; Scott, J. N.; Lambert, G. G.; Salih, A.; Shaner, N. C.; Hughes, T. E., Local Electric Field Controls Fluorescence Quantum Yield of Red and Far-Red Fluorescent Proteins. *Front Mol Biosci* **2021,** *8*, 633217.
12. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **2004,** *22* (12), 1567-72.
13. Shcherbo, D.; Murphy, C. S.; Ermakova, G. V.; Solovieva, E. A.; Chepurnykh, T. V.; Shcheglov, A. S.; Verkhusha, V. V.; Pletnev, V. Z.; Hazelwood, K. L.; Roche, P. M.; Lukyanov, S.; Zaraisky, A. G.; Davidson, M. W.; Chudakov, D. M., Far-red fluorescent tags for protein imaging in living tissues. *Biochem J* **2009,** *418* (3), 567-74.
14. Lam, A. J.; St-Pierre, F.; Gong, Y.; Marshall, J. D.; Cranfill, P. J.; Baird, M. A.; McKeown, M. R.; Wiedenmann, J.; Davidson, M. W.; Schnitzer, M. J.; Tsien, R. Y.; Lin, M. Z., Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat Methods* **2012,** *9* (10), 1005-12.

15. Pandelieva, A. T.; Baran, M. J.; Calderini, G. F.; McCann, J. L.; Tremblay, V.; Sarvan, S.; Davey, J. A.; Couture, J. F.; Chica, R. A., Brighter Red Fluorescent Proteins by Rational Design of Triple-Decker Motif. *ACS Chem Biol* **2016,** *11* (2), 508-17.
16. Chu, J.; Haynes, R. D.; Corbel, S. Y.; Li, P.; González-González, E.; Burg, J. S.; Ataie, N. J.; Lam, A. J.; Cranfill, P. J.; Baird, M. A.; Davidson, M. W.; Ng, H. L.; Garcia, K. C.; Contag, C. H.; Shen, K.; Blau, H. M.; Lin, M. Z., Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nat Methods* **2014,** *11* (5), 572-8.
17. Li, Y.; Palmer, A. G., Narrowing of protein NMR spectral lines broadened by chemical exchange. *J Am Chem Soc* **2010,** *132* (26), 8856-7.
18. Griffin, J. M.; Ashbrook, S. E., Solid-state NMR of high-pressure silicates in the earth's mantle. *Annu Rep NMR Spectrosc* **2013,** *79*, 241-332.
19. Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L., Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc Natl Acad Sci U S A* **2010,** *107* (47), 20257-62.
20. Moore, M. M.; Oteng-Pabi, S. K.; Pandelieva, A. T.; Mayo, S. L.; Chica, R. A., Recovery of red fluorescent protein chromophore maturation deficiency through rational design. *PloS one* **2012,** *7* (12), e52463.
21. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-47.
22. Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D., A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007,** *450* (7171), 913-6.
23. Alford, S. C.; Ding, Y.; Simmen, T.; Campbell, R. E., Dimerization-dependent green and yellow fluorescent proteins. *ACS Synth Biol* **2012,** *1* (12), 569-75.
24. Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **2002,** *99* (12), 7877-82.
25. Hockenberry, A. J.; Wilke, C. O., Evolutionary couplings detect side-chain interactions. *PeerJ* **2019,** *7*, e7280.
26. Mosavi, L. K.; Peng, Z. Y., Structure-based substitutions for increased solubility of a designed protein. *Protein Eng* **2003,** *16* (10), 739-45.
27. Carballo-Amador, M. A.; McKenzie, E. A.; Dickson, A. J.; Warwicker, J., Surface patches on recombinant erythropoietin predict protein solubility: engineering proteins to minimise aggregation. *BMC Biotechnol* **2019,** *19* (1), 26.
28. Chan, P.; Curtis, R. A.; Warwicker, J., Soluble expression of proteins correlates with a lack of positively-charged surface. *Sci Rep* **2013,** *3*, 3333.
29. Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Section 1.3, Chemical bonds in biochemistry*. W H Freeman: New York, 2002.
30. Islam, M. M.; Kobayashi, K.; Kidokoro, S. I.; Kuroda, Y., Hydrophobic surface residues can stabilize a protein through improved water-protein interactions. *FEBS J* **2019,** *286* (20), 4122-4134.
31. Malleshappa Gowder, S.; Chatterjee, J.; Chaudhuri, T.; Paul, K., Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *ScientificWorldJournal* **2014,** *2014*, 971258.

Chapter 5: Materials and Methods

Computational protein design of Superdecker RFPs. All calculations were performed with the Triad protein design software (Protabit, Pasadena, CA, USA) using a Monte Carlo with simulated annealing search algorithm for rotamer optimization. The 63-member backbone ensemble of mRojo-SD1 prepared as described above was used to provide templates for computational design. Following extraction of protein heavy-atom coordinates for the highest occupancy conformer, hydrogen atoms were added using the *addH.py* application in Triad, and the resulting structures were energy-minimized by performing 100 steps of conjugate gradient energy minimization in Cartesian space using the Phoenix energy function (see below) with added covalent terms from the Dreiding force field.[1]

The 2002 backbone-dependent Dunbrack rotamer library[2] with expansions of ± 1 standard deviation around $\chi 1$ and $\chi 2$ was used to provide amino-acid side-chain conformations to be threaded onto each backbone template. Chromophore rotamers were generated by allowing all 36 combinations of the following dihedral angles around the indicated bonds: CA2-CB2-CG2-CD1 (–15°, –10°, –5°, 0°, +5°, and +10°), N2-CA2-CB2-CG2 (–5°, 0°, +5°, +10°, +15°, and +20°). Side-chain rotamers were optimized on each backbone template using the following amino acids (numbering based on the mRojo-SD1 crystal structure): residue 99 (F, I, L, M), 143 (D, S), 161 (F, I, L, M), 177 (F, I, L, M, V), 165 (F, I, L, M), and 199 (L, M). Side-chain rotamers of second-shell residues S62, H63, E97, R163, and Y197 were also optimized. The searched sequence space thus consisted of 1,280 sequences. To ensure that the lowest energy rotamer configuration was generated for each sequence, 1280 individual single-state design trajectories (one per sequence) were performed on each template using the *cleanSequences.py* app within Triad.

Sequences were scored on each backbone template using the Phoenix energy function,[3, 4] which consists of a Lennard-Jones 12–6 van der Waals term from the Dreiding II force field[1] with atomic radii scaled by 0.95, a direction-dependent hydrogen bond term with a well depth of 8.0 kcal mol^{-1} and an equilibrium donor-acceptor distance of 2.8 Å,[5] an electrostatic energy term modelled using Coulomb's law with a distance-dependent dielectric of 10, an occlusion-based solvation potential with scale factors of 0.05 for nonpolar burial, 2.5 for nonpolar exposure, and 1.0 for polar burial,[3] and a secondary structural propensity term with amino acid phi-psi propensities derived from the method of Shortle.[6]

Following design, the top 8 ranked sequences across the ensemble that were predicted to adopt the superdecker structure as their most stable conformation were experimentally characterized.

Computational protein design of mSandys. All calculations were performed with the Triad protein design software (Protabit, Pasadena, CA, USA). The structure of mRojoA was obtained from the Protein Data Bank (PDB code: 3NEZ[7]). Following extraction of protein heavy-atom coordinates for the highest occupancy conformer, hydrogen atoms were added and the resulting structure was energy-minimized by performing 100 steps of conjugate gradient energy minimization in Cartesian space using the Phoenix energy function (see below) with added covalent terms from the Dreiding force field.[8] Individual chains of the minimized mRojoA structure were extracted to generate an ensemble of 4 backbone templates for multistate design.

Multistate design using a Monte Carlo with simulated annealing search algorithm for rotamer optimization and a fitness function corresponding to the Boltzmann weighted average energy (Temperature = 300 K) for sequence scoring was performed using the *multi_design.py* app within Triad. The 2002 backbone-dependent Dunbrack rotamer library[9] with expansions of ± 1 standard deviation around $\chi 1$ and $\chi 2$ was used to provide amino-acid side-chain conformations to be threaded onto each backbone template. Chromophore rotamers were generated by allowing all 36 combinations of the following dihedral angles around the indicated bonds: CA2-CB2-CG2-CD1 (–15°, –10°, –5°, 0°, +5°, and +10°), N2-CA2-CB2-CG2 (–5°, 0°, +5°, +10°, +15°, and +20°). Side-chain rotamers of first-shell residues surrounding the phenolate moiety of the chromophore (**Supplementary Figure S3.1a**) were optimized on each backbone template using the following amino acids: residue 63 (I, L, V), 143 (I, L, S, V), 161 (I, L, M, V), 163 (I, L, M, V), 197 (I, L, V), and 199 (I, L, M, V). Side-chain rotamers of second-shell residues S62, K70, M97, L165, V177, and E148 were also optimized, and the T16V mutation was introduced. The searched sequence space thus consisted of 2,304 sequences. Sequences were scored on each backbone template using the Phoenix energy function,[7, 10] which consists of a Lennard-Jones 12–6 van der Waals term from the Dreiding II force field[8] with atomic radii scaled by 0.95, a direction-dependent hydrogen bond term with a well depth of 8.0 kcal mol^{-1} and an equilibrium donor-acceptor distance of 2.8 Å,[11] an electrostatic energy term modelled using Coulomb's law with a distance-dependent dielectric of 10, an occlusion-based solvation potential with scale factors of 0.05 for nonpolar burial, 2.5 for nonpolar exposure, and 1.0 for polar burial,[7] and a secondary structural propensity term

with amino acid phi-psi propensities derived from the method of Shortle.[12] The top 10 ranked sequences (**Supplementary Figure S3.1b**) were experimentally characterized.

Red fluorescent protein genes. Amino-acid sequences for all RFPs described here are listed on **Table 5.1**. His-tagged (N-terminus) genes for mRojoA and mCherry cloned into pET-11a (Novagen) were a gift from Stephen Mayo.[7] Codon-optimized and his-tagged (N-terminus) genes for mRojoA variants, mRojo-SD1 variants, mCherryIAG and mCherryIQLGV cloned into the pET-29b(+) vector (Novagen) via *Nde*I and *Xho*I were obtained from Twist Bioscience. All plasmids were transformed into *E. coli* BL21-Gold (DE3) cells for protein expression. The mSandy2 gene was prepared as described in the *Random mutagenesis* section below.

Random mutagenesis. Three rounds of random mutagenesis, starting from the mSandy1 gene and then on the pooled RFP genes from the top 1% cells obtained after cell sorting (see below), were performed using the protocol described by McCullum *et al.*[13] Briefly, four cycles of error-prone polymerase chain reaction using *Taq* DNA Polymerase (5 U) in 1X Standard *Taq* Buffer pH 8.3 (New England Biolabs) supplemented with $MnCl_2$ (0.5 mM) and a mixture of deoxynucleotides (1 mM dTTP, 1 mM dCTP, 0.2 mM dATP and 0.2 mM dGTP), were used to introduce approximately two mutations on the 732-bp RFP open reading frame. The resulting mutant library was cloned into pET-11a (Novagen) via *Nde*I and *Bam*HI. Plasmids were transformed into *E. cloni* 10G Elite electrocompetent cells (Lucigen) and plated onto lysogeny broth (LB) agar supplemented with 100 µg/mL ampicillin. Following incubation at 37°C overnight, all colonies on the agar plate were

collected using 1 mL of LB supplemented with 100 µg/mL ampicillin, and the plasmid pool was harvested by DNA extraction (E.Z.N.A. Plasmid DNA Mini Kit, Omega Bio-tek). Purified plasmid DNA was then transformed into electrocompetent *E. coli* BL21-Gold (DE3) cells (Agilent) for protein expression.

Combinatorial site-saturation mutagenesis of mCherry surface patch libraries. Combinatorial libraries were generated by site-saturation mutagenesis by overlap extension polymerase chain reaction[14] using NNS degenerate codons and mCherry as the starting template. Libraries were subcloned into pET-11a (Novagen) via *Nde*I and *Bam*HI restriction sites and the plasmids were transformed into *E. cloni* 10G ELITE cells (Lucigen). Introduction of NNS codons at the desired positions in the libraries was verified by DNA sequencing. Plasmids were then transformed into *E. coli* BL21-Gold (DE3) cells for protein expression.

Fluorescence-activated cell sorting. BL21-Gold (DE3) cells containing the mSandy1 random library or the mCherry surface patch libraries obtained as described above were used to inoculate 100 mL of LB supplemented with 100 µg/mL ampicillin. Cells were grown at 37°C with shaking until they reached an optical density of 0.6 at 600 nm. At this point, cells expressing random mSandy1 mutants were harvested by centrifugation. Cells were washed twice with sterile phosphate buffer (20 mM sodium phosphate buffer, pH 7.4, 50 mM sodium chloride). Pellets were stored at 4°C to allow for chromophore maturation. After two days, cells were resuspended in sterile phosphate buffer to a concentration of 10^7 cells/mL, and filtered using 40-µm cell strainers (Falcon).

Cells were sorted (top 1%) using a MoFlo Astrios EQ cell sorter (Beckman Coulter) according to fluorescence intensity (λ_{ex} = 561 nm, λ_{em} = 614 nm) into sterile flow cytometry test tubes (Beckman Coulter) containing LB supplemented with 100 µg/mL ampicillin. Sorted cells were washed twice with supplemented LB and grown overnight in 50-mL cultures at 37°C with shaking. 5 mL of this culture was used for extraction of plasmid DNA. Two additional rounds of random mutagenesis and cell sorting were performed using the procedures described above.

Manipulation of sorted cells of the mSandy1 library. Cells collected from the third round of sorting were used to inoculate LB agar plates supplemented with 100 µg/mL ampicillin. Following overnight incubation at 37°C, 96 individual colonies were transferred onto a Nunc MicroWell 96-well polypropylene plate (Thermo Scientific) containing 250 µL of LB supplemented with 100 µg/mL ampicillin. This plate was covered with a sterile breathable rayon film membrane (VWR) and incubated overnight at 37°C with shaking. Cells included in this "mother" plate were used to inoculate cultures for protein expression and purification as described below.

Manipulation of sorted cells of the mSandy1 library. Cells collected from the third round of sorting were sorted into Nunc MicroWell 96-well polypropylene plates (Nalgene) (two plates per library) containing 250 µL of LB broth supplemented with 100 µg/mL ampicillin. Plates were covered with sterile breathable rayon film membranes (VWR) and incubated overnight at 37°C with shaking.

Protein expression and purification in microplates. BL21-Gold (DE3) cells transformed with plasmids containing computationally designed or sorted RFP variants were used to inoculate deep 24-well plates (Whatman) containing 5 mL of Overnight Express Instant TB medium (Novagen) supplemented with 100 µg/mL ampicillin (pET11-a) or kanamycin (pET-29b(+)). These plates were sealed with sterile and pierced silicone sealing mats (Axygen) and incubated overnight at 37°C with shaking. Cells were then harvested by centrifugation and washed twice with phosphate buffer pH 7.4. Pellets were stored at 4°C to allow for chromophore maturation. After two days, cell pellets were resuspended in 400 µL of lysis buffer (50 mM sodium phosphate buffer, pH 8.0, 300 mM sodium chloride, 2.5 mM imidazole, 1X Bugbuster protein extraction reagent [Novagen], 1 mg/mL hen egg white lysozyme [Sigma], 10 U benzonase nuclease [Novagen]) and plates were incubated at 25°C with gentle shaking for 20 minutes. After incubation, clarified lysates were collected by centrifugation and proteins were purified using HisPur Ni-NTA spin plates (Thermo Scientific) according to the manufacturer's protocol. Protein purity was verified by SDS-PAGE.

Protein expression and purification in large batches. RFP variants were expressed in *E. coli* BL21-Gold (DE3) cells (Agilent) using LB supplemented with 100 µg/mL ampicillin (pET11-a) or 50 µg/mL (pET29b(+)). Cells were grown at 37°C with shaking until they reached an optical density at 600 nm of 0.6. Protein expression was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (Thermo Scientific) and cultures were incubated overnight at 16°C with shaking. Cells were then pelleted by centrifugation and resuspended in 8 mL lysis buffer (100 mM potassium phosphate buffer, pH 7.4, 5 mM

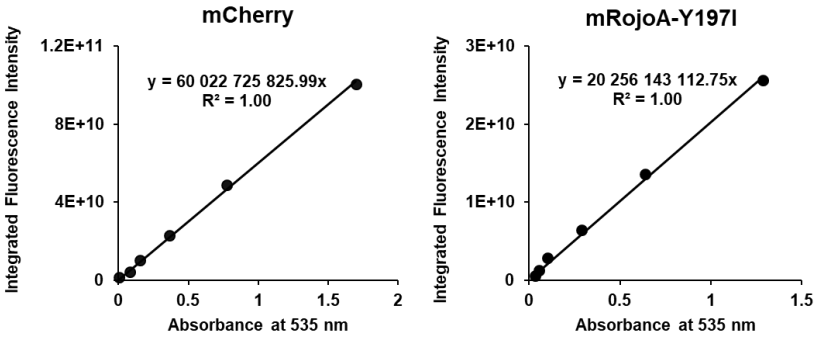
imidazole, 1 mg/mL lysozyme, 50 U benzonase nuclease [Novagen]). Cells were lysed using an EmulsiFlex-B15 cell disruptor (Avestin) and lysates were harvested by centrifugation. Proteins were purified by Ni-NTA affinity chromatography using Econo-Pac chromatography columns (Bio-Rad) according to the manufacturer's protocol. Eluted fractions were desalted using Macrosep Advance centrifugal devices (Pall) into phosphate buffer solution. Protein samples were stored at 4°C for two days to allow for chromophore maturation prior to spectroscopic characterization. For crystallography experiments, an additional purification step consisting of gel filtration into 20 mM sodium phosphate buffer pH 7.4 was performed using an ÄKTA pure (GE Healthcare) fast protein liquid chromatography system equipped with a Superdex 75 (GE Healthcare) column.

Protein quantification. Absorbance of 1:10 purified protein:8 M guanidine hydrochloride in a quartz cuvette was measured at 280 nm using a Spectramax Plus 384 Microplate Reader (Molecular Devices) at room temperature. Protein concentration was calculated using the Beer's law equation, where c is the protein concentration (in M), A is the absorbance at 280 nm, 10 is the dilution factor, ε is the extinction coefficient of the protein as obtained from the ExPASy ProtParam online tool (Swiss Institute of Bioinformatics) (in $M^{-1}\,cm^{-1}$), and l is the path length of the cuvette of 1 cm:

$$c = \frac{A \times 10}{\varepsilon l}$$

Spectroscopic characterization. Absorption, excitation and emission spectra were measured in 1X phosphate-buffered saline (137 mM sodium chloride, 2.7 mM potassium

chloride, 10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4) or in sodium phosphate buffer pH 9 (), using an Infinite M1000 (Tecan) or Spectramax i3 (Molecular Devices) plate reader at room temperature. Quantum yields were extrapolated by comparing the integrated fluorescence intensity of the mutant proteins with that of equally absorbing samples of mCherry and mRaspberry (quantum yields of 0.23 and 0.15,[15] respectively) with excitation at 535 nm. An example is shown below:



$$\Phi_{mRojoA\text{-}Y197I} = (\Phi_{mCherry\ (standard)})(y_{mRojoA\text{-}Y197I})/y_{mCherry}$$

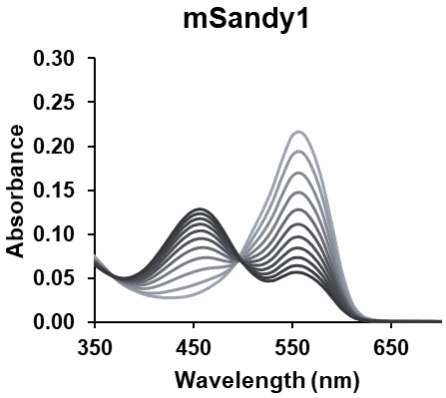
$$\Phi_{mRojoA\text{-}Y197I} = (0.22)(20\ 256\ 143\ 112.75)/60\ 022\ 725\ 825.99$$

$$\Phi_{mRojoA\text{-}Y197I} = 0.074$$

Measurement of the quantum yield is not impacted by non-fluorescent impurities in the sample since only the red-chromophore-containing molecules contribute to the absorbance and emission values (respectively 535 nm and 560-700 nm) used in the calculation.

Extinction coefficients (ε) were determined using the method described by Kredel and colleagues,[16] which takes into account the presence of green chromophore in the population of RFP molecules, giving a more accurate estimate. Briefly, protein samples were diluted (1:10) into Britton-Robinson buffers[17] pH 11–13 to allow measureable slow denaturation of the native red chromophore (peak at 568–605 nm, depending on the RFP) to the green form (452 nm). From the absorbance spectrum recorded at different time

intervals, a ratio of extinction coefficients of the native chromophore species to that of the denatured species (44,000 M^{-1} cm^{-1}) was calculated, from which the value for the red chromophore was extrapolated. Absorbance spectra were also measured at pH 7.4, pH 11 and pH 13 to correct for absorbance peak variation under denaturing and non-denaturing conditions at different pH. A general example is shown below, where light-to-dark curves represent red-to-green chromophore denaturation overtime:



$$\varepsilon_{\text{red chromophore}} = 44{,}000\ M^{-1}\ cm^{-1} \times \frac{\Delta A_{\text{red peak (pH 11–13)}}}{\Delta A_{\text{green peak (pH 11–13)}}} \times \frac{A_{\text{red peak (pH 7.4)}}}{A_{\text{red peak (pH 11)}}} \times \frac{A_{\text{green peak (pH 13)}}}{A_{\text{green peak (pH 11–13)}}}$$

Similarly to the quantum yield measurement, protein impurities should not impact the extinction coefficient. This is because the method developed by Kredel and colleagues[16] strictly measures the absorbance of the red and denatured green chromophores at their respective wavelength (respectively around 550-600 nm and around 455 nm). This avoids underestimation of the extinction coefficient that could be due to presence of non-fluorescent impurities.

pH profiles and pK_a measurements. Proteins were diluted to a final concentration of 0.5 – 1.0 µM into a series of Britton-Robinson buffers titrated with sodium hydroxide at various

pH (2 – 12). Full emission spectra at different pH were measured in black 96-well microplates (Greiner Bio-One) with excitation at the appropriate wavelength (505 – 525 nm, depending on the RFP). pH profiles were obtained by plotting the integrated fluorescence as a function of pH. pK_a values were extrapolated using a sigmoidal curve (1 apparent pK_a) or a biphasic curve (2 apparent pK_a's) equations, solved using the least squares method until F* = F for each point :

$$F^* = \frac{H_1 - H_0}{1 + 10^{(pKa - pH)n}} + H_0 \ ,$$

$$\text{or} \qquad F^* = \frac{H_1 - H_0}{1 + 10^{(pKa_1 - pH)n_1}} + \frac{H_2 - H_1 - H_0}{1 + 10^{(pKa_2 - pH)n_2}} + H_0$$

where *F* is the normalized integrated fluorescence intensity at a given pH, H_0 is the baseline, H_1 the first plateau, H_2 the second plateau, and n_1 and n_2 the Hill coefficients.

Protein crystallization of mRojo-SD1. Purified mRojo-SD1 was concentrated using Amicon Ultra-15 3K centrifugal devices (Millipore Sigma) to a concentration of 34 mg mL^{-1} in 200 mM sodium phosphate buffer pH 7.4. Crystallization drops were prepared by mixing 1 µL of protein solution with 1 µL of the mother liquor and sealed inside a reservoir containing an additional 500 µL of the mother liquor solution. Crystals used for X-ray data collection were obtained from mother liquor containing 0.1 M Tris buffer (pH 8.5), 0.1 M sodium chloride, and 20% PEG-3350 (**Supplementary Table S2.1**).

Protein crystallization of mSandy2. Purified mSandy2 was concentrated using Amicon Ultra-15 3K centrifugal devices (Millipore Sigma) to a concentration of 34 mg mL^{-1} in 200 mM sodium phosphate buffer pH 7.4. Crystallization drops were prepared by mixing 1 µL

of protein solution with 1 µL of the mother liquor and sealed inside a reservoir containing an additional 500 µL of the mother liquor solution. Crystals used for X-ray data collection were obtained from mother liquor containing 0.1 M Tris buffer (pH 8.5), 0.1 M sodium chloride, and 20% PEG-3350 (**Table 3.2**).

X-ray data collection and processing. Prior to X-ray data collection, crystals were harvested and cryoprotected by brief soaking in a 1:1 mixture of crystallization mother liquor and 4.0 M trimethylamine-*N*-oxide, then flash-cooled by rapid plunging into liquid nitrogen. We collected single-crystal X-ray diffraction data on beamline 8.3.1 at the Advanced Light Source. The beamline was equipped with a Pilatus3 S 6M detector (Dectris), the X-ray energy was set to 11111 keV, and the crystals were maintained at a cryogenic temperature (100 K) throughout the course of data collection.

We processed the X-ray data using the Xia2 software,[18] which performed indexing, integration, and scaling with XDS and XSCALE,[19] followed by merging with Pointless.[20] A resolution cutoff (2.05 Å) was taken where the signal-to-noise ratio of the individual reflection measurements fell to an average value of 1.0.

Structure determination. We obtained initial phase information for calculation of electron density maps by molecular replacement using the program Phaser,[21] as implemented in v1.17.1.3660 of the PHENIX suite.[22] We identified eight copies of the protein in the asymmetric unit using the coordinates from mCherry (PDB ID: 2H5Q[4]), consistent with an

analysis of Matthews probabilities for the observed unit cell and molecular weight of the protein.[23, 24]

Next, we rebuilt the initial model using the electron density maps calculated from molecular replacement. We then performed additional, iterative refinement of atomic positions, individual atomic displacement parameters (B-factors), and occupancies, until the model reached convergence. All model building was performed using Coot 0.8.9.2[25] and refinement steps were performed with phenix.refine (v1.17.1.3660) within the PHENIX suite.[22, 25] Restraints for the red and green chromophores were generated using phenix.elbow,[26] starting from coordinates available in the Protein Data Bank[27] (PDB ligand ID: NRQ or CH6 for red or green chromophore, respectively). Further information regarding model building and refinement of mSandy2 is presented in **Table 3.2**, and that of mRojo-SD1 is presented in **Supplementary Table S2.1**. In the case of mRojo-SD1, a time-averaged ensemble was generated with phenix.ensemble_refinement implemented in PHENIX.[28] To prepare the structures for ensemble refinement, low-occupancy conformers were removed, and occupancies adjusted to 100% using phenix.pdbtools. Hydrogen atoms were then added using phenix.ready_set. This procedure yielded an ensemble containing 63 templates.

Molecular dynamics. All simulations were performed using the Amber 2018 software with the ff14sb force field[29]. Long-range electrostatics (>8 Å) were modeled using the particle mesh Ewald method[30], and the SHAKE[31, 32] algorithm was used to treat all bonds involving hydrogens as constraints, allowing a time step of 2 fs. Heavy atom coordinates of the major conformer from chain A were extracted from the computational model of each *Superdecker*

RFP generated by Triad. Chromophore parameters were generated within the Antechamber module. Following coordinate extraction, hydrogen atoms were added using Reduce,[33] and the resulting protein molecule was placed in an octahedral box with periodic boundary conditions where the distance between the protein surface and the box edges was set to 10 Å. After addition of explicit OPC[34] water molecules, charges on protein atoms were neutralized with Na^+ and Cl^- counter-ions. Next, we performed energy minimization of the molecular system in two stages. In the first stage, we minimized water molecules and ions while keeping the protein fixed using 1000 steps of steepest descent followed by 1000 steps of conjugate gradient. In the second stage, we minimized the entire system using 2000 steps of conjugate gradient. Then the system was heated from 0 to 300 K using the Langevin thermostat.[35] The system was equilibrated under an NVT ensemble for 1 ns at a temperature of 300 K using the Berendsen coupling algorithm[36] to maintain constant temperature. A second equilibration step under an NPT ensemble was performed for 1 ns with a constant pressure of 1 bar and a temperature of 300 K, using the Berendsen barostat[36]. Following removal of the position restraints, a 200-ns production run under the Langevin thermostat was initiated from the final snapshot of the NPT equilibration PMEMD (Duke RE, Pedersen LG (2003) PMEMD 3 University of North Carolina - Chapel Hill). At the end of the simulation, 1000 snapshots separated by 0.2 ns along the production trajectory were extracted for further analysis.

Table 5.1. Amino-acid sequences of red fluorescent proteins.

Protein	Sequence
mRojo-VHSV	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSHQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGSEASSERMYPEDGALKGEIKVRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNANYKLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
Mutant4.3	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSHQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGSEASSER**L**YPEDGALKGEIKVR**Y**KLKDGGHYDAEVKT TYKAKKPVQLPGAY**Y**ANYKLDITSHNEDYTIVEQYER**S**EGRHSTGGMDELYK
mRojo-SD1	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSHQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERV**E**NFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGSEASSER**L**YPEDGALKGEIK**R**R**Y**KLKDGGHYDAEVKT TYKAKKPVQLPGAY**Y**ANYK**M**DITSHNEDYTIVEQYER**S**EGRHSTGGMDELYK
mSD1.0	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSHQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERV**E**NFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMGSEASSER**L**YPEDGALKGE**M**K**R**RLKLKDGGHYDAE**I**KT TYKAKKPVQLPGAY**Y**ANYKLDITSHNEDYTIVEQYER**S**EGRHSTGGMDELYK
mSD1.1	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSHQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERV**E**NFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMGSEASSER**L**YPEDGALKGEIK**R**RLKLKDGGHYDAE**M**KT TYKAKKPVQLPGAY**Y**ANYKLDITSHNEDYTIVEQYER**S**EGRHSTGGMDELYK
mSD1.5	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSHQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERV**E**NFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMGSEASSER**L**YPEDGALKGE**L**K**R**RLKLKDGGHYDAE**M**KT TYKAKKPVQLPGAY**Y**ANYKLDITSHNEDYTIVEQYER**S**EGRHSTGGMDELYK
mRojoA	MGHHHHHHGVSKGEEDNMAIIKEFMRFKTHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNANYKLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mRojo-Y197I	MGHHHHHHGVSKGEEDNMAIIKEFMRFKTHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.1	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**KLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.7	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK

mSandy0.9	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy1	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**KLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy2	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**R**F**KLKDGGHY**V**AEVKT TYKAKKPVQLPGAY**Y**AN**I**K**M**DITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy1- L163V/L199M	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**K**M**DITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mCherry	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIK**Q**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYN**V**N**I**KLDITSHNEDYTIVEQYER**A**EGRHSTGGMDELYK
mCherry	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK
mCherryA	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNV**A**IKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK
mCherryQLV	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKT TYKAKKPVQLPGAY**Q**V**L**IKLDITSHNEDYTIVEQYERAEG**V**HSTGGMDELYK
mCherryIG	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKKTMGWE**I**SSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNVNI**G**LDITSHNEDYTIVEQYERAEGRHSTGGMDELYK
mCherryIAG	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKKTMGWE**I**SSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNV**A**I**G**LDITSHNEDYTIVEQYERAEGRHSTGGMDELYK

mCherryIQLGV	MGHHHHHHGVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKKTMGWE**I**SSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKT TYKAKKPVQLPGAY**Q**V**L**I**G**LDITSHNEDYTIVEQYERAEG**V**HSTGGMDELYK

Mutations from mRojo-VHSV, from mRojoA or from mCherry are underlined and highlighted in bold.
The sequences corresponding to their respective chapter are separated by thick lines in the table.

References of Materials and Methods

1. Wannier, T. M.; Moore, M. M.; Mou, Y.; Mayo, S. L., Computational Design of the β-Sheet Surface of a Red Fluorescent Protein Allows Control of Protein Oligomerization. *PloS one* **2015,** *10* (6), e0130582.
2. Wannier, T. M.; Gillespie, S. K.; Hutchins, N.; McIsaac, R. S.; Wu, S. Y.; Shen, Y.; Campbell, R. E.; Brown, K. S.; Mayo, S. L., Monomerization of far-red fluorescent proteins. *Proc Natl Acad Sci U S A* **2018,** *115* (48), E11294-E11301.
3. Ash, C.; Dubec, M.; Donne, K.; Bashford, T., Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers in medical science* **2017,** *32* (8), 1909-1918.
4. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-47.
5. Lambert, T. J., FPbase: a community-editable fluorescent protein database. *Nat Methods* **2019,** *16* (4), 277-278.
6. Moore, M. M.; Oteng-Pabi, S. K.; Pandelieva, A. T.; Mayo, S. L.; Chica, R. A., Recovery of red fluorescent protein chromophore maturation deficiency through rational design. *PloS one* **2012,** *7* (12), e52463.
7. Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L., Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc Natl Acad Sci U S A* **2010,** *107* (47), 20257-62.
8. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., Dreiding - a Generic Force-Field for Molecular Simulations. *J. Phys. Chem.* **1990,** *94* (26), 8897-8909.
9. Dunbrack, R. L., Jr., Rotamer libraries in the 21st century. *Curr Opin Struct Biol* **2002,** *12* (4), 431-40.
10. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., Iterative approach to computational enzyme design. *Proc Natl Acad Sci U S A* **2012,** *109* (10), 3790-5.
11. Dahiyat, B. I.; Mayo, S. L., Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **1997,** *94* (19), 10172-7.
12. Shortle, D.; Stites, W. E.; Meeker, A. K., Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* **1990,** *29* (35), 8033-41.
13. McCullum, E. O.; Williams, B. A.; Zhang, J.; Chaput, J. C., Random mutagenesis by error-prone PCR. *Methods Mol Biol* **2010,** *634*, 103-9.
14. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R., Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **1989,** *77* (1), 51-9.
15. Wang, L.; Jackson, W. C.; Steinbach, P. A.; Tsien, R. Y., Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A* **2004,** *101* (48), 16745-9.
16. Kredel, S.; Nienhaus, K.; Oswald, F.; Wolff, M.; Ivanchenko, S.; Cymer, F.; Jeromin, A.; Michel, F. J.; Spindler, K.-D.; Heilker, R.; Nienhaus, G. U.; Wiedenmann, J., Optimized and Far-Red-Emitting Variants of Fluorescent Protein eqFP611. *Chemistry & Biology* **2008,** *15* (3), 224-233.
17. Britton, H. T. S.; Robinson, R. A., CXCVIII.-Universal buffer solutions and the dissociation constant of veronal. *Journal of the Chemical Society (Resumed)* **1931,** (0), 1456-1462.

18. Winter, G., xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **2010,** *43* (1), 186-190.
19. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.
20. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 1), 72-82.
21. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007,** *40* (Pt 4), 658-674.
22. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 213-21.
23. Kantardjieff, K. A.; Rupp, B., Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **2003,** *12* (9), 1865-71.
24. Weichenberger, C. X.; Rupp, B., Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallogr D Biol Crystallogr* **2014,** *70* (Pt 6), 1579-88.
25. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 352-67.
26. Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D., electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **2009,** *65* (Pt 10), 1074-80.
27. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M., The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem* **1977,** *80* (2), 319-24.
28. Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzhanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A., Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nature communications* **2020,** *11* (1), 4808.
29. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015,** *11* (8), 3696-3713.
30. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.
31. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327-341.
32. Miyamoto, S.; Kollman, P. A., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **1992,** *13* (8), 952-962.

33. Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **1999,** *285* (4), 1735-47.
34. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *The journal of physical chemistry letters* **2014,** *5* (21), 3863-3871.
35. Davidchack, R. L.; Handel, R.; Tretyakov, M. V., Langevin thermostat for rigid body dynamics. *J Chem Phys* **2009,** *130* (23), 234101.
36. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

Chapter 6: Summary and Outlook

In this thesis, we investigated three different rational design approaches to increase RFP brightness. In the first project, we aimed to rigidify the RFP chromophore and thereby increase its quantum yield via the creation of a superdecker motif of aromatic rings that includes the chromophore phenolate moiety. Although we still need to confirm the presence of this structural motif by X-ray crystallography, our results demonstrate that this approach yields enhanced brightness, albeit at higher pH values than standard physiological conditions. We therefore concluded that improving brightness through the design of a superdecker motif to rigidify the chromophore can be challenging and results in only modest increases in brightness, prompting us to explore alternative approaches.

In the second project, we investigated a different approach to rigidify the RFP chromophore, using aliphatic instead of aromatic residues, which we postulated would provide better combinations of side-chain rotamers to pack tightly around the chromophore phenolate moiety. Using this aliphatic packing strategy, we successfully obtained a large brightness increase, resulting in an approximately 10-fold increase in quantum yield. We also demonstrated the benefit of combining this methodology with directed evolution, which further enhanced brightness and yielded the brightest monomeric *Discosoma*-derived RFP with an emission maximum above 600 nm reported to date. However, compared to the parent protein, we observed a large hypsochromic shift in emission wavelength. These undesirable changes to the emission wavelength were also observed in the superdecker RFPs. In both cases, mutations were introduced directly next to the chromophore, likely causing changes to its electronic environment and thereby emission wavelength, an undesired side-effect.

In order to eliminate undesirable impacts on other spectroscopic properties of the chromophore, in the third and last project, we attempted to introduce mutations far from the chromophore to improve brightness without changing its electrostatic environment and therefore emission wavelength. Based on previous NMR results, we identified clusters of surface residues whose structural rigidity was found to correlate with quantum yield. We hypothesized that mutagenesis of these clusters could result in mutation combinations that would cause further rigidification of the protein and thereby enhance brightness. However, mutagenesis of these clusters of residues did not result in the identification of RFP mutants displaying enhanced brightness. Nevertheless, the mutations also did not negatively impact any of the other spectroscopic properties of the chromophore, suggesting that identifying beneficial combinations of mutations that exclusively increase brightness could be possible. Future work should therefore focus on targeting additional clusters of residues for which correlations between structural dynamics and brightness were observed. Combining this approach with aliphatic packing, as applied in the previous project, and a more direct experimental screening method such as high-throughput fluorescence lifetime imaging could be highly beneficial. Such methodologies would enable to increase side chain packing to rigidify regions of the protein structure where rigidity and brightness were found to be correlated, and more efficiently screen for mutants with enhanced quantum yield and brightness.

In this thesis, we demonstrated that different rational design approaches can be used to increase RFP brightness, some of which are more successful than others, and have the potential to be used as general strategies to create useful RFPs for different biological applications.